Adaptive Preferences
and
Children’s Options

Rosa Elizabeth Terlazzo

A thesis submitted for the degree of
Doctor of Philosophy
of
The Australian National University

May 2013
Statement of Originality
The work contained in this thesis is my own and, unless otherwise attributed, is not the result of joint effort or research. No material in this thesis has been previously submitted for the award of a degree at any institution of higher learning.

Rosa Elizabeth Terlazzo
Abstract

What should we think of the woman who believes that it is proper for her husband to make all of the family’s decisions? How do we respond to the mother who thinks that her domestic abuse is deserved? One popular explanation of such cases is that people’s preferences may “adapt” to harmful or unacceptable circumstances that they see as unavoidable, so that they come to endorse states of affairs that they should rightly resist. Yet the concept of adaptive preferences is highly controversial: while its defenders argue that it is a useful tool for understanding the way in which members of marginalized groups can acquiesce to their own oppression and marginalization, its detractors argue that employing the concept unavoidably and disrespectfully treats adults as if they were children who do not know what is good for them. But at the same time, the concept is incredibly under-theorized: defenders and detractors of adaptive preferences alike rarely offer a clear account of the concept against which their claims can be tested.

In this thesis, I propose and defend an autonomy-based account of adaptive preferences with four aims: a) conceptual clarity, b) the ability to show respect for persons, d) conduciveness to the political project of reducing marginalization, and d) recognition of and attention to the differential needs of children and adults. To achieve the first aim, I outline what I call an “indirect substantive” account of autonomy. This account uses substantive content indirectly to determine whether a person’s preferences count as autonomous. Because my account of autonomy uses substantive content indirectly, it allows the possibility that any preference could count as autonomous, and therefore non-adaptive. In this way, it shows respect for persons by recognizing the role that their own processes of moral reasoning play in determining their good. But by using substantive content, I also ensure that the concept can aid the political project. In order for a preference to count as autonomous, and therefore non-adaptive, I argue that the person developing it must have been exposed to alternatives to the preference developed that were both live and valuable. In this way, while people may legitimately prefer options that seem to third parties to be unattractive or marginalizing, these preferences only deserve full deference when they have been chosen from among valuable alternatives. Finally, I turn to the issue of children. This issue is especially important, since previous discussions of adaptive preferences have failed to recognize that the needs of adults who have already developed adaptive preferences are very different from the needs of children in the process of developing them. I argue for ways of operationalizing my theoretical account of adaptive preferences that can both show respect for the already-formed preferences of adults (including those that are adaptive), and prevent children from forming adaptive preferences in the first place.
# Table of Contents

Abstract .......................................................................................................................... v
Table of Contents ........................................................................................................... vii
Acknowledgements ....................................................................................................... ix

## Introduction

What do I mean by “adaptive”? .................................................................................. 5
What do I mean by “preference”? ............................................................................... 7
What do I mean by “an account of adaptive preferences”? .................................... 11
Who has adaptive preferences? ............................................................................... 12
Who should respond to adaptive preferences? ...................................................... 14
An overview of the thesis ......................................................................................... 15

## Chapter One: Perfectionist Accounts of Adaptive Preference

Perfectionism ................................................................................................................. 17
Perfectionist accounts of adaptive preferences: benefits ...................................... 18
Nussbaum’s account of adaptive preferences ......................................................... 20
Perfectionist accounts of adaptive preferences: drawbacks ................................... 24
Transitioning from perfectionism to political liberalism ...................................... 32
Conclusion .................................................................................................................. 41

## Chapter Two: Deliberative Accounts of Adaptive Preference

Deliberative projects .................................................................................................... 43
Benefits of deliberative response: efficacy, agency, autonomy ............................. 45
Benefits of broadly deliberative accounts: provision of secondary recognition respect and maintenance of standards for moral reasoning ............. 49
Two theoretical weaknesses: inability to respect faith-based moral reasoning, and indeterminacy of the point of success ................................................. 51
An empirical weakness: deliberation may not root out adaptiveness in practice .......... 55
Implicit bias .................................................................................................................. 59
Problems for adaptive preference ........................................................................... 64
Conclusion .................................................................................................................. 70

## Chapter Three: Children and Adaptive Preferences

Adaptive preferences and primary recognition respect ......................................... 71
Autonomy and children ............................................................................................. 74
Establishing a difference between children and adults ......................................... 82
A Dilemma .................................................................................................................... 90
Conclusion .................................................................................................................. 92
Acknowledgements

While all omissions and errors that remain in this thesis are my own, there would surely have been far more of them without the involvement, support, and constructive criticism of many. Hardly a week went by without a conversation with a faculty member or fellow PhD student that helped me to rethink and improve my thesis, and I am grateful for every one of these. A few people, however, deserve my special thanks.

Among these, the members of my panel are first and foremost. Every interaction that I have had with Thomas Pogge has made my work stronger and my skin thicker, and I am a better philosopher for both. Christian Barry encouraged me in the early days of my thesis to explore outside projects that excited me, and it was ultimately one of these that led to the subject on which I settled. The consistent support and enthusiasm that Marilyn Friedman showed for my project was invaluable, as was her willingness to join my panel even as her affiliation with my program was ending. And finally, Scott Wisor has spent more hours discussing more disparate issues in my thesis and philosophy more generally in more depth and on shorter notice than any early (or late) career academic could ever be expected to. All four have my sincere gratitude.

More broadly, my time at the Research School of Philosophy, and at the Centre for Applied Philosophy and Public Ethics before that, has exposed me to more bright and committed philosophers than I can count. In particular, the idea that grew into my thesis has its roots in collaborative work done with Luara Ferracioli, and my interactions with her, both across an office and across oceans, have always helped to remind me both of what matters in the world and of why I love philosophy. Adrian Currie has always been ready with comments both serious and humorous, and both have been equally appreciated. Seth Lazar and Jonathan Quong have both pushed me on important points and made valuable suggestions, without which my work would have been much weaker. The whole FemPov team (and especially Kieran Donaghye) has provided me with the most supportive, encouraging, and exciting introduction to interdisciplinary research that I could have hoped for. And although unaffiliated with the ANU, Serene Khader has been one of the most significant influences on my work. I eagerly anticipate further chances both to agree and disagree with her in the future.

And of course, my thanks, as always, go to Jon Herington. He makes it all better: my philosophy, my life, the world.
Introduction

At its most basic, the idea of adaptive preferences is fairly straightforward: when a person's option set is limited, her preferences may adapt to her situation in such a way that she comes to prefer something that she otherwise would not. And at this basic level, the idea is also uncontroversial: everyone has a limited option set, and everyone develops at least some preferences that they would not have developed had their option set been different. A person who grows up close to the sea might understandably be more likely to enjoy sailing than a person in a landlocked state. A person who excels at sports but struggles with academics might understandably be more likely to prefer a physically demanding career than a person who is academically gifted but lacks physical prowess. Our circumstances make us familiar with some options rather than others, cast some options rather than others as desirable and worth pursuing, and place obstacles in the way of some options while facilitating the pursuit of others. All of this, so far, seems like common sense.

When the term "adaptive preference" comes up in the academic literature, however, it is often used in a more normative way: it suggests the more controversial idea that those preferences that are to be counted as "adaptive" should be treated somehow differently from other, "nonadaptive" preferences. Here, the term tends to pick out a much narrower set of preferences: those that, in addition to having been developed in limited circumstances, are bad for us, or inauthentic, or liable to social suspicion, or somehow not meaningfully our own. In other words, preferences that are adaptive in the normative sense are ones that, in some important sense, ought not to have been developed. And accordingly, they ought not to have the same force as other preferences in determining the way that the world ought to be.

But how do we make sense of such ideas? As Serene Khader rightly notes, the literature that addresses the subject is of surprisingly little help. In her recent book *Adaptive Preferences and Women's Empowerment*, she calls attention to the fact that discussions of adaptive preferences are often vague and rely on examples that seem intuitively to be adaptive rather than providing an account of the conditions that cause preferences to count as adaptive in the normative sense.¹ Theorists who use the term

---

point, for instance, to women who continue to bear children to the detriment of their own health, women who do not question the abuse that they receive from their husbands, women who uncomplainingly take on physically taxing work that men would not be asked to do, groups of people in objectively ill health who nevertheless have very low rates of self-reported morbidity, and sweatshop employees who do not protest the exploitative circumstances of their employment. Among these theorists, there seems to be a general sense that we will know adaptive preferences when we see them. But Khader notes further that reliance on such intuitive examples lumps preferences of very different sorts together under the term "adaptive": preferences formed consciously and unconsciously, preferences that exist because of a lack of options and those that exist because of a lack of awareness, preferences that undermine a person's whole sense of self, and those that do not. The vagueness combined with the sheer variety of the examples, then, makes it difficult for a reader to determine just what it is that we talk about when we talk about adaptive preferences.

In part, we can excuse this lack of clarity by considering the projects of the theorists who utilize the concept. Generally, these theorists are focused on another goal in the achievement of which the concept of adaptive preferences plays only a small part. Amartya Sen, one of the two most prominent theorists to draw on the concept, is primarily concerned with establishing his capability approach as a metric of advantage superior to the utilitarian metric of welfare. He relies on the concept of adaptive preference only to illustrate the shortcomings of this latter metric. Martha Nussbaum, the other prominent theorist to use the concept, uses it in aid of her claim that social justice requires a substantive list of entitlements owed to human beings everywhere. Without such a list, she argues, adaptive preferences will remain a hazard. Other theorists also use the concept in a supporting manner. Ann Levey and Ann E. Cudd both reference the concept as part of a larger discussion about the compatibility of

---

9 Ibid.
14 Khader's own account of adaptive preferences offers a notable exception to this problem of vagueness. Indeed, her account of the concept is the clearest and the most comprehensive in the literature. It has, however, some significant drawbacks, which I discuss in Chapter One.
feminism and liberalism; Naila Kabeer brings the concept up as one of the factors that makes it difficult to measure women's empowerment; Anita Superson uses it to illustrate the inadequacies of informed desire tests; and Susan Moller Okin uses the concept to show that her Rawlsian-inspired approach to thinking about justice is superior to the "interactive feminist" alternative. For theorists involved in such dissimilar projects, the know-it-when-you-see-it method of talking about adaptive preferences might be sufficient, since any preference that does strike the reader as both adaptive and problematic will serve to pump the reader's intuitions in favor of the argument being presented. For their immediate purposes, then, the task of establishing that adaptive preferences exist can indeed take precedence over establishing what they are.

I argue, however, that there are two important reasons that even these theorists should want a clearer account of adaptive preferences. The first relates to the clarity of the debate. This debate is enormously controversial, and the concept of adaptive preferences has as many detractors as it has advocates. Among the common criticisms of the concept are the claims that it encourages seeing those said to have adaptive preferences as patients rather than agents, that it licenses denying those with adaptive preferences the right to act according to their preferences, that it entails seeing persons' preferences as dismissible, that using it allows more powerful actors to substitute their own voices and judgment for those of less powerful and more vulnerable actors, and that it encourages outsiders to see those with adaptive preferences as irrational, victimized, or psychologically damaged. The first reason to want a clearer account of adaptive preferences is simply to determine whether these

14 Susan Moller Okin (1994) "Gender Equality and Cultural Differences" Political Theory, 22(1).
18 For something like this criticism, see Alison Jaggar (2006) "Reasoning About Well-being: Nussbaum's Method of Justifying the Capabilities Approach" Journal of Political Philosophy, 14(3), esp. p. 317. Note that in this article, Jaggar does not direct this criticism at the concept of adaptive preferences directly, but rather at what she calls Nussbaum's "non-platonist substantive-good approach" to justifying the capabilities. That approach, however, is motivated in large part by the problem of adaptive preferences.
criticisms actually apply. Whether they are warranted will depend on the exact nature of the account given, and until we have such an account, critics and champions of the concept will simply be speaking past one another.

The second reason to want a clear account of adaptive preferences relates as much to practice as to theory. The theorists who use the concept in the normative sense are social and political theorists, and their intuitive discussions of the concept, in Khader's words, aim "to supply a widely applicable narrative about why actual [persons] perpetuate their own oppression."20 This narrative, however, is not an end in itself. The intuitions that adaptive preference theorists draw from their readers are intuitions about the real-world problems of social inequality, injustice, and domination. They rightly see these problems as urgent ones in need of attention, and their work has, at least in part, the political agenda of finding real solutions to those problems.

The work done on adaptive preferences is in large part motivated by this important and practical political project: the project of identifying the preferences that spring from these problematic circumstances, and of proposing and enacting responses to those preferences that can help to empower persons to desire and aim for more in their lives than their unjust and inequitable circumstances seem to offer. But the criticisms raised are real and important ones, and it would make a mockery of this political project if attempts to alleviate important social ills required disrespectful treatment of the very individuals harmed by them. So a clearer account of adaptive preferences is needed not only to determine whether these criticisms apply, but also to ensure that they do not. If it were indeed the case that the concept of adaptive preferences necessarily warranted these criticisms, then they would constitute good grounds for rejecting the use of the concept. But if it can be shown that the concept is compatible with the respectful treatment of persons, not only in the way that it conceptualizes adaptive preferences, but also in the way in which it identifies and responds to them, then it does valuable work for the political project and should therefore be maintained.

Three major goals of my thesis, then, are already implicit in the literature on the subject and need only be made explicit: I will aim to develop an account of adaptive preferences that 1) provides conceptual clarity, 2) furthers the political project with

20 Serene J. Khader (2012) "Must Theorizing about Adaptive Preferences Deny Women's Agency?" Journal of Applied Philosophy, 29(4), p. 303. Khader locates this project directly within the feminist literature, and so focuses on explaining why women perpetuate their oppression. Many adaptive preferences will indeed belong to women, since women are more likely to be marginalized and oppressed than their male counterparts. For reasons explained below, however, I think it important to recognize that the problem is not uniquely a women's problem. Accordingly, I refer to "persons" rather than "women".
which the concept is associated, and 3) does so in a way that makes possible the respectful treatment of persons. But I will have a fourth major goal as well, which has as yet received no attention in the literature: using that account to do justice to the needs of children. All of the criticisms raised against the concept involve the needs of adults, and the ways in which adults may respectfully be treated. But while adults with adaptive preferences certainly have an interest in being treated respectfully, children also have a strong interest in not forming adaptive preferences in the first place. As I will argue in Chapter Three, a tension can arise between these two sets of interests, and an acceptable account of adaptive preferences must balance the two. But because no discussion of the needs of children has yet been undertaken in the literature, the interests of adults have tended to be prioritized. My thesis attempts to right this imbalance.

The thesis is comprised of two halves. In the first half, I develop a set of considerations that should inform a desirable account of adaptive preferences. In the second half, I develop this account. Before I offer an overview of the thesis, however, I want to clarify several further points.

What do I mean by “adaptive”?
The term “adaptive” refers to the genesis of a preference, and as I noted above, nearly every preference that a person could potentially have will be adaptive in the non-normative sense. But while I recognize the non-normative use of the term as a legitimate one, I want to establish a convention for the use of the term “adaptive” in this thesis. Some theorists defend the concept of adaptive preference precisely because they use it in the non-normative way, and some theorists who are primarily concerned with the normative issues of inequality and social injustice nevertheless reserve the term “adaptive preference” for use in the non-normative sense and utilize modifiers or other conventions to distinguish between instances of problematic and non-problematic adaptive preferences. I, however, take the political project mentioned above to be the primary reason for having the concept in the first place, and

---

21 For instance, Donald Bruckner is primarily concerned with the ways in which adaptive preferences can be rational and beneficial, and treats those preferences that arise in conditions of injustice and inequality as only one small subclass of adaptive preferences. Donald W. Bruckner (2009) “In Defense of Adaptive Preferences” Philosophical Studies, 142 (3).

22 For instance, Serene Khader uses the phrase “inappropriately adaptive preferences” to refer to preferences that are adaptive in what I have been calling the normative sense, while Ann Levey relies primarily on context to distinguish between normative and non-normative instances of adaptive preferences.
I tend to find the proposed modifiers bulky and unnecessarily confusing. Accordingly, when I use the term “adaptive” in this thesis, I will always (unless otherwise explicitly noted) intend it in the normative sense in which it means a preference which is liable to social suspicion on the grounds of having been formed in accordance with avoidable social inequality or domination.\(^{23}\) When I use the colloquial term “we” in this thesis, I refer to academics and other interested parties who share the political agenda to which this normative and political use of the term is attached; I hope, however, that even those theorists who are primarily concerned with the non-normative use of the term will find something of interest in my thesis.

Using the term in this normative sense will of course have implications for the substance of the account that I eventually develop. And while the intuitive examples that other theorists have used to talk about adaptive preferences are not sufficient for a full account of the concept, they are in some sense necessary. The political project that I take on is concerned with the ways in which social injustice, social inequality, and domination can influence preferences, and if the preferences that an account identifies as adaptive have nothing to do with this political project, or if those preferences identified as nonadaptive seem to have clear links to these social conditions, then the account will be of something quite different than the concept that we are trying to get at. Of course, any normative account of adaptive preferences will at some points challenge our intuitions, and this should be expected. If the account did not occasionally identify an intuitively adaptive preference as in fact nonadaptive, or vice versa, then it would be superfluous. We could simply dispense with it and rely on our intuitions. What a desirable account of adaptive preferences must do is refine and justify the work done by intuition about domination and social injustice, rather than simply replicating it.

Accordingly, I offer the first consideration intended to inform the development of my own account of adaptive preferences:

**Consideration 1: Appropriate scope.** An account of adaptive preferences should reliably identify preferences clearly related to the political project as adaptive and preferences clearly unrelated to the political project as nonadaptive.

As evidenced by my adoption of the political agenda, I take seriously the idea that persons' preferences can be deeply influenced by unjust or dominating social conditions.

\(^{23}\) Note that I am not committed to the stronger normative use of the term suggested by other theorists where “adaptive” means something more like “irrational”, “inescapably bad”, or “always in need of remedy”. These issues will be taken up in detail in the body of the thesis.
circumstances. I do not, however, think that persons' circumstances determine their preferences. Often, people will naturally come to prefer the practices and values that are shared by those who surround them as they grow up, and this is why it is problematic for persons to develop preferences in circumstances where their options are limited to bad ones. But persons may also develop preferences in response to norms rather than in accordance with them, so that exposure to a norm causes a person to reject a practice or value more forcefully than she would if she had lacked exposure. The problem is that this kind of rejection all too often requires a heroic individual with exceptional independence of mind – and the ability to escape limiting circumstances should not depend upon being exceptional in these ways.

I also do not take adaptiveness to imply irrationality. When a person may be unable to escape her social circumstances, acquiescing will often be easier and more psychologically comfortable than fighting those circumstances, and it is not irrational to prefer psychological comfort to difficulty and uncertainty of success. If a person truly cannot avoid her circumstances then perhaps her adaptive preferences should indeed be presumed to be normatively authoritative – psychological comfort is certainly to be preferred to incessant pining after a genuinely impossible alternative. But life is virtually never such that a person's circumstances are truly unavoidable. Nearly always, socially created circumstances could be different, given enough hard work on the part of enough people. The normative authority of adaptive preferences should be questioned not because they are necessarily irrational to hold, but because conscious disappointment with an avoidable status quo is the first step towards changing that status quo for the better.

What do I mean by "preference"?
The term "preference" gets used in many different ways. It may mean that a person endorses some value across the board, regardless of the alternatives to which it is compared. In this sense, a person might have a preference for equality as a good which makes any situation better. "Preference" might also refer to a three place relation where one thing is preferred to another thing by some particular person. In such a case, we can say that Rosa prefers spring to summer, but this tells us nothing about her preferences about any other season or subject. A preference might also refer to an

---

24 In this thesis, I leave aside metaphysical questions about determinism and free will.
all-things-considered preference that takes feasibility of options into account. Here, Tony might prefer to be a high school teacher given that he does not have the talent to be a successful athlete, while still recognizing that he would be happier as a basketball star if he could become one. Alternatively, a preference could be for one option among many where feasibility plays no role at all, as it might if Tony retained his preference for being a basketball player despite the limitations. The term preference might also be used to refer only to desires exclusively about one's own interests, where one's feelings about the wider world are counted as desires of some other sort. Alternatively, one can hold that preferences can include desires about the world beyond oneself, so that it makes sense to say that John prefers that Cindy wear her hair long. Most broadly, the term may be used to refer to all of these relationships between a person and her judgments, including her tastes, her own perceived interests, her positions on values like equality and justice, and her willingness to allow feasibility to serve as a limiting feature of her judgments. 26

Some readers may be frustrated by the fact that I adopt the last of these possibilities. And indeed, those readers would not be alone: one of the primary objections to the use of the concept of adaptive preferences hinges on the idea that the term “preference” ought to have only one use. The objection takes this general form: the concept of “adaptive preferences” assumes that persons actually desire things that are bad for them. But persons do not in fact desire things that are bad for them; rather, they acquiesce to circumstances that they realize are bad for them in cases where they are powerless to challenge those circumstances, or in cases where they receive some other pay-off for acquiescing. Since people do not actually desire the circumstances that are bad for them, the argument concludes, people do not have adaptive preferences, and we ought to stop speaking and writing as if they do. 27


There is some merit to this argument. The literature is indeed full of cases in which persons who seem to endorse disadvantaging practices actually speak critically of those practices in private, or undertake covert acts of resistance. So if the concept of adaptive preference did refer only to wholehearted endorsement of disadvantaging practices, I would generally agree with the critics who argue that the concept is not a useful one.28

It is not clear to me, however, why we should use the concept in this way. If we look at the actual instances in which the term is used, very few refer to cases of fully endorsed disadvantaging practices. Sen, for instance, often points to examples in which persons enjoy small mercies in otherwise oppressive lives,29 or those in which persons fail to see their situations as dismal because they have literally no experience of what a better situation might be like.30 Susan Moller Okin explains the persistence of practices like foot binding and clitoridectomy as instances in which women recognize the agony that accompanies the practices, but “perpetuate the cruelties, inflicting them, or at least allowing them to be inflicted on their daughters” because they recognize the necessity of these practices for life within their societies.31 And when Martha Nussbaum talks about adaptive preferences, she notes that women often refuse to undertake organized opposition to the status quo not because they think that the status quo is desirable, but rather because “women are afraid that change will make things worse.”32

Since those concerned about adaptive preferences are concerned with “preferences” of all sorts, the previous objection that persons rarely actually desire their own disadvantage provides no reason to jettison the concept of adaptive preferences all together. Instead, the objection offers a reason to expand the set of the sorts of preferences that an account of adaptive preferences must be concerned with in order to be useful. While many of these instances of adaptive preferences lack the full endorsement of the individuals involved, they all involve cases in which individuals make choices or accept situations that disadvantage them in some way, and do so as a

28 I say that I would “generally” agree because I continue to think that there exist cases in which persons do genuinely desire things that are bad for them as a result of the circumstances in which their preferences formed. Indeed, I can point to several such instances in myself. Rather than agreeing that such cases do not exist, then, I agree only that cases of different sorts are too often misidentified as this kind of genuine desire.


30 Sen (2002) "Health: Perception versus Observation".


result of their limited circumstances. And since the concept of adaptive preferences is a
useful one for getting at this class of instances, it should be retained and refined in
order to have that vital discussion in a clearer way.

Consider women in the developing world who place less value on their own
education or the education of their daughters than they do on the education of their
sons.33 Or consider women with professional ambitions in the developed world who
decide to stay at home to care for their young children and allow their husbands to
continue to work.34 Both of these choices may be rational ones: in the developing
world, female children are often married off and play no role in supporting their
parents in old age, so their education may seem like a waste of money to mothers with
limited resources; in the developed world, men have higher earning power than
women, and when one parent must stay home with children, it often makes economic
sense for a man rather than a woman to remain in the workforce. The list of
potentially rational reasons for failing to (in at least some sense) prefer one’s own good
goes on. Disadvantaged individuals may not value oppression, but may nevertheless
value the social roles that support that oppression.35 Or they might object to
oppression, but have developed sincere tastes for some of the aesthetic values that are
correlated with it.36 Other individuals might be tasked with the care of community
members who rely on them, like children, the disabled, or the very old, and might
therefore give attention to the needs of these voiceless parties to the detriment of their
own interests.37 But while all of these choices are at least partially rational ones, they
are formed by conditions that doubtless seem inevitable, and they serve to
disadvantage not only the specific individuals who make them, but all others who are
constrained by the same norms that these choices support. And I contend that while
we should not treat these choices as illegitimate, we should also refuse to treat them as
unproblematic. Retaining the category of adaptive preference allows us to walk this
fine line. Accordingly, I propose a second consideration:

34 Cordelia Fine (2010) Delusions of Gender: How Our Minds, Society, and Neurosexism Create Difference,
36 Ellen Gruenbaum (2006) “Sexuality Issues in the Movement to Abolish Female Genital Cutting in
Sudan” Medical Anthropology Quarterly, 20(1). Consider as well women in the West who wish to
achieve common standards of thinness and beauty even if they object to the way in which these
norms help to objectify women.
Routledge.
Consideration 2: Variety of preferences. An account of adaptive preferences must be concerned with preferences at all levels, from those that are deeply and sincerely held to those that represent compromises with unfavorable realities.

What do I mean by “an account of adaptive preferences”? An exhaustive discussion of adaptive preferences will have three parts: concept, identification, and response. A conceptual account of adaptive preferences must be able to offer a set of conditions according to which a preference qualifies as adaptive in theory. This is a significant task, and offering such an account will be the primary aim of my thesis. The political project, however, is attached not only to understanding adaptive preferences, but to ameliorating them, and because I am committed to this agenda, I will also consider two other issues: how adaptive preferences can be identified in practice, and how they can be responded to. Since both of these issues will be affected significantly by context, my discussions in these areas will be more preliminary. But as I mentioned above, showing that the concept can treat persons respectfully at each step is one of the main goals of my thesis, so these preliminary discussions are important and should be the site of further research.

Dividing an account of adaptive preferences into these three distinct elements – concept, identification, and response – is, admittedly, somewhat artificial. Theorists who engage with the concept often have a focus that lends itself to one of these categories, yet their discussion also manages to cast some light on issues in the other two. Even Khader’s account, by far the clearest in the literature, often deals simultaneously with issues that fall into several of these categories. And indeed, this should not be surprising: in many cases, the same task or concept will serve more than one function. An account of adaptive preferences might use the act of deliberating both to identify which preferences are adaptive, and to develop a set of responses to any adaptive preferences identified. Alternatively, it might utilize the idea of exposure to alternatives at all three stages: surviving exposure to alternatives could serve to conceptually establish the nonadaptiveness of a preference, while the same exposure in practice could simultaneously help to identify which actual preferences were adaptive and act as a response to the existence of adaptive preferences by offering persons with such preferences new alternatives to pursue.

Nevertheless, such division is useful for developing a desirable account of adaptive preferences. Because objections can be leveled at any of the three elements, distinguishing between them will help to ensure that my account succeeds at each
level. Secondly, making such a distinction will help to clarify the intuitive accounts of the other theorists who use the concept. Since they tend to use the concept without discussing it in depth, they often focus far more on one element than the others. By pinpointing the element that receives the most attention in each of their accounts, I can make explicit note when I extrapolate the other elements of their accounts in order to develop the considerations that will influence my own.

Who has adaptive preferences?

In the broad sense, we all have adaptive preferences. The normative sense in which I use the term, however, is limited to preferences that spring from socially disadvantaging circumstances. It should, then, come as no surprise that the persons who are going to have adaptive preferences on my account will be those who suffer social disadvantage. Some readers will doubtless object to my account on these grounds, arguing that an account that attributes adaptive preferences only to the disadvantaged widens the already overly-broad gulf between the advantaged and the disadvantaged. Alison Jaggar, for instance, seems to have this worry in mind when she writes that

[R]aising questions of false consciousness only with respect to non-Western women who defend their cultures could be read as suggesting that these women's moral perceptions are less reliable than the perceptions of Western women, whose consciousness is supposedly higher or truer.38

Worries of this sort are important ones. The political project that attends the normative use of the term is motivated by concerns about domination and social injustice, and any use of the concept of adaptive preferences that exacerbates either inequality or injustice should be forcefully resisted by those who share this project. There are, however, at least two reasons that associating adaptive preferences with disadvantage need not raise such worries.

First, persons are not disadvantaged simpliciter. Rather, they are disadvantaged in some particular way, and in relationship to some particular part of their identity. A black man in the United States may be disadvantaged by virtue of his race, but advantaged by virtue of his gender. A poor Brahman in India may be advantaged in virtue of his caste, but disadvantaged economically. And indeed, even those features that are often taken to confer advantage can also confer disadvantage in

more particular circumstances: during the Vietnam War, young American men were certainly placed at a disadvantage by their gender, which made them eligible for conscription into the armed forces. In a less formal way, gender norms in many parts of the world continue to make men more common targets of many kinds of violence than women. 39 If we understand disadvantage in this selective way, then attaching the concept of adaptive preference to disadvantage need not serve to create two classes of persons, or widen the gulf between already established classes. Instead, it can do the opposite: it can help to draw attention to disadvantage wherever it exists, both in the places where we would expect to find it and in the places in which we are least accustomed to considering it.

Second, although the concept of adaptive preference has often carried connotations of poor moral reasoning or inauthenticity of preference, I will argue throughout this thesis that such connotations are contingent and can and should be excised. Reserving the term “adaptive” for preferences associated with disadvantage need not be coupled with the suggestion that only persons with adaptive preferences misidentify the good. To the contrary, persons in all social situations can have preferences that are bad for them, and these preferences can form without any exposure to disadvantage. And in the same way, this reservation of the term “adaptive” need not be coupled with the implication that adaptive preferences are thoughtless or irrational, while nonadaptive preferences are rational and considered. Indeed, those with adaptive preferences, forced as they are to confront some of life’s most pressing difficulties, may consider their preferences and their options much more carefully than those who have the luxury of choosing between many desirable opportunities. So while I incorporate and justify some of the connotations that come with the concept of adaptive preferences, like their association with disadvantage, I reject other connotations, like those of irrationality and inauthenticity, that exacerbate existing social inequalities and render the concept unable to show respect for persons identified as having adaptive preferences.

Given the tie between adaptive preference and disadvantage, it should be unsurprising that much of the discussion of adaptive preference occurs in the literature on feminist philosophy and development ethics. Although they are not the only groups to suffer social disadvantage, women and the poor in the developing world certainly

suffer more than their fair share of it, and the theorists who write on issues that pertain to them are understandably particularly concerned with such disadvantage. Accordingly, issues of gender and poverty will arise frequently in this thesis. However, the frequency with which they will be discussed should not be taken as an indication that only women and the poor develop adaptive preferences. Rather, my discussion can and should be extended to those in the developed world as well.

Who should respond to adaptive preferences?

I have said that the primary purpose of my thesis is to develop a desirable conceptual account of adaptive preferences, and my audience for this task will primarily be philosophers and other academics. I have said as well, however, that I will engage with questions of identification and response, and while such questions should be of interest to the theorists whom I intend as my primary audience, I take it that these theorists will not themselves be the actors best situated to respond to actual adaptive preferences in practice. Before I move on to the body of my thesis, then, I want to briefly discuss the actors that I take to more appropriately fill this role.

Any actor whose actions may influence the context in which preferences are formed could potentially find themselves in the position to respond to adaptive preferences. Actors, however will vary both according to the frequency with which they encounter circumstances of social inequality, and according to the extent to which they have a mandate to act on those circumstances. The greater the frequency of exposure and the mandate of an actor, the greater should be the responsibility of the actor to make an effort to respond to adaptive preferences that they might encounter, and to do so in permissible and considered ways.

Frequency and mandate will come together most obviously in the case of the state. States administer domestic programs that impact upon social inequality and injustice through both national and local government structures, and the way in which these programs are designed and implemented cannot help but have an impact on the

---

40 The importance of cases like these explains what may have struck the reader as a startling omission: as of yet, there has been no note of Jon Elster's foundational account of adaptive preferences. This may be especially surprising since, until Khader's account of the concept, Elster's was the most extensive account available in the literature. Elster's account, however, will receive little attention in this thesis because it counts as adaptive only those non-autonomous preferences that replace previously held preferences -- it cannot deal with the kinds of preferences, like many of those that are most striking in the cases of women and the poor, that arise from norms that people grow up with since birth. For Elster's account, see Jon Elster (1982) "Sour grapes -- utilitarianism and the genesis of wants" in Amartya Sen and Bernard Williams (eds.), *Utilitarianism and Beyond*, Cambridge: Cambridge University Press; Jon Elster (1983) *Sour Grapes: Studies in the Subversion of Rationality*, Cambridge: Cambridge University Press.
way in which preferences held in those circumstances can be formed and satisfied. In the developing world, however, tasks that would normally be undertaken by states often fall to nongovernmental organizations, and the work that they do can also have a profound effect on the circumstances that influence persons' preferences. When I discuss responses to adaptive preferences in this thesis, I will focus primarily on these sorts of state and non-state actors. It should be noted, however, that other actors — among them, parents, friends, private employers, and community groups — may also have reason to be concerned about adaptive preferences and reason to attempt responses to them. While much of what I say may apply to them, they will not be my main focus. In interpersonal situations, a separate set of questions arises, and while these questions are important ones, a full treatment of them would require extensive discussion that lies beyond the scope of this thesis. They represent, however, interesting areas for further research.

An overview of the thesis

In the first half of the thesis, I develop considerations to inform my account by focusing on three topics: perfectionist accounts of adaptive preferences, deliberative accounts of adaptive preferences, and children.

In Chapter One, I discuss what I call perfectionist accounts of adaptive preferences, which use substantive perfectionist content to identify the preferences that count as adaptive in theory. I argue that while such accounts provide a plausible relationship between the concept and the political project associated with it, they fail to show persons two important kinds of respect.

In Chapter Two, I discuss what I call deliberative accounts of adaptive preferences, which count as adaptive those preferences that persons would reject after reflective deliberation. I argue that, while these accounts rightly treat individuals as the ultimate authorities on their own good, they fail to recognize that deliberation may not always be sufficient to alter preferences when genuine alternatives to the status quo seem remote.

In Chapter Three, I discuss children and adaptive preferences. I argue that the preferences and behaviors of adults in a child's community constrain the genuine alternatives that the child may consider for herself, but that adults may also develop genuine interests in particular adaptive preferences by living in accordance with them for a long enough period of time. Recognizing this, I argue that a justifiable account of the concept must succeed both in respecting the adults who have already developed
adaptive preferences, and in meeting the needs of children who have an interest in developing non-adaptive preferences – and that these two goals can come into tension with one another.

In the second half of the thesis, I develop an account of adaptive preferences in response to these considerations. I defend an autonomy-based account of adaptive preferences, since such an account succeeds in treating persons as the ultimate authorities on their own good and recommends responses aimed at increasing autonomy rather than simply changing behavior.

In Chapter Four, I show that conceiving of adaptive preferences as autonomy deficits need not open persons up to coercive paternalistic intervention. Indeed, I show that conceiving of autonomy in the correct way actually provides reasons in favor of allowing persons to make their own choices even in cases of adaptive preferences.

In Chapter Five, I develop an account of autonomy that both succeeds in treating persons as the ultimate authorities on their own good, and that appropriately identifies the conditions under which preferences should count as adaptive. This account of autonomy is what I call an "indirect substantive" account, which uses substantive content indirectly to determine a preference's adaptiveness. I argue that my account can show persons a form of respect that responds to their interest in being recognized as competent moral reasoners.

In Chapter Six, I discuss the identification of adaptive preferences in the real world, and argue that different sets of considerations hold for children and adults. Since adults already have a history with even their adaptive preferences, it will be difficult to identify the substantive content that can appropriately be used in the account of autonomy developed in the previous chapter. Since children do not yet have these commitments, however, valuable substantive content will be easier to identify. I argue that the job of identifying this substantive content should be left to "multi-sited" critics who have a range of experiences, including in the local context in question.

In Chapter Seven, I discuss responses to adaptive preferences. I show how the interests of children can justify the use of incentives targeted at adults. These incentives can be permissibly aimed at adults in order to expand the set of options that a child sees in her immediate context. I note, however, that the use of incentives can sometimes be coercive, and propose a provisional set of considerations for responding to this problem.
Chapter One: Perfectionist Accounts of Adaptive Preference

Although one need not look far to find references to adaptive preferences in the human development literature, clear accounts of the concept are remarkably rare. But despite this scarcity, some of the major accounts resemble each other in important ways. Two major accounts – Martha Nussbaum’s in *Women and Human Development* and Serene J. Khader’s in *Adaptive Preferences and Women’s Empowerment* – are importantly similar in two ways: both rely on perfectionist conceptions of the good, but both utilize thinner conceptions of the good than perfectionism often employs. In this chapter I examine the reasons that perfectionism might offer a compelling grounding for an account of adaptive preferences, but ultimately argue that there are significant problems with understanding adaptive preferences in relationship to a perfectionist account of human good. In light of both these benefits and these drawbacks, I begin to develop a set of considerations to be applied in the development of my own account of adaptive preferences.

Perfectionism

Broadly, perfectionist moral theories hold that there is an objective human good at which persons ought to aim.\(^1\) According to Thomas Hurka, a perfectionist theory claims that the good life consists in developing to a high degree whichever properties it picks out as those that “constitute human nature or are definitive of humanity – [that] make humans humans.”\(^2\) The candidates for this role are many, but any perfectionist moral theory must identify some specific good or set of goods that it directs persons to pursue. These goods will count as goods whether or not people actually want to pursue them, since perfectionism hold that “what is good for its own sake for a person is fixed independently of her attitudes and opinions towards it.”\(^3\) And from perfectionist moral theories, it is a short leap to perfectionist political theories.


that hold that it is the role of the state to pursue and develop these goods, whether or not people endorse them.

Of course, there is little agreement on which particular excellences are relevant, and the content of particular perfectionist theories is therefore often controversial. Recognizing that this controversy becomes particularly problematic in the political realm, many contemporary political philosophers have rejected state support of perfectionist values. Following John Rawls, they recognize that in liberal societies where freedom of conscience is respected, reasonable people will disagree about the appropriate ends of human life and the human excellences that correspond to them. Accordingly, they hold that it is illegitimate for the state to promote one set of excellences over others, encouraging its development in the lives of citizens. In Martha Nussbaum’s words, the state should not “be in the business of ranking and ordering comprehensive doctrines ... by saying that these are genuinely valuable and those are not.” This is because, she claims, promoting one conception of the good will necessarily put those who endorse other conceptions of the good at a disadvantage. And since comprehensive doctrines “are so deeply a part of people’s search for the meaning of life... public denigration of those doctrines puts those people at a disadvantage, suggesting that they are less worthy than other citizens.” Since there is such broad disagreement about the good, political liberals like Nussbaum and Rawls worry that endorsing a perfectionist set of values cannot help but disrespect the many citizens who think that the excellences of human life lie elsewhere.

**Perfectionist accounts of adaptive preferences: benefits**

Serene Khader offers the clearest and most plausible perfectionist account of adaptive preferences in the literature. On her account, in order to be adaptive, a preference must meet three conditions:

---


46 Ibid, p. 22.
An adaptive preference is a preference that (1) is inconsistent with a person’s basic flourishing, (2) was formed under conditions nonconducive to her basic flourishing, and (3) that we do not think a person would have formed under conditions conducive to basic flourishing.\textsuperscript{47}

Khader makes her account explicitly perfectionist by placing the concept of flourishing at the heart of her first condition for adaptiveness: a preference is adaptive (in part) insofar as it fails to serve its holder’s basic flourishing.\textsuperscript{48} The modifier “basic”, however, is important – Khader, like Rawls, recognizes the great diversity of persons’ ideas about the good, and the accompanying difficulty of settling on a set of human excellences that persons can reach agreement upon. Therefore, Khader insists that her perfectionist account only take a stand on the things that constitute human flourishing at the most basic level, where she holds that disagreements about the good will be rarer.\textsuperscript{49} And to further ensure that the perfectionist list that informs her account is not a controversial one, she refuses to suggest the content of that list herself. Instead, she holds that such a perfectionist list must be the result of an actual cross-cultural dialogue that allows people from many different backgrounds to have their voices heard.\textsuperscript{50} In this way, she avoids one of the central problems of most perfectionist accounts: since her account of human flourishing takes no stand on the excellences of human nature, those with many different accounts of truly excellent human lives can agree upon the minimal list of basic goods that she proposes. In this way, although a perfectionist doctrine is proposed, no comprehensive doctrine is privileged over others.

Such a limited perfectionist account has at least two \textit{prima facie} advantages as a basis for an account of adaptive preferences. The first is conceptual: it draws the line between adaptive and nonadaptive preferences in a way that accords with the political project discussed in the Introduction. An account that identifies a millionaire’s preference for caviar as adaptive treats adaptiveness as unrelated to either social injustice or domination. A perfectionist account like Khader’s, on the other hand, identifies what it is that is unjust. If justice requires the equal (or sufficient) distribution of \textit{something}, then the opportunity for basic flourishing is an extremely plausible candidate. By identifying the elements of basic human flourishing, we can


\textsuperscript{48} Like me, Khader uses the term “preferences” loosely. Different kinds of preferences can be inconsistent with flourishing in different ways: a preference “revealed” by behaviour may directly compromise flourishing; a higher-order endorsement of a value might fail to appreciate the value of basic flourishing without directly necessarily compromising it.

\textsuperscript{49} \textit{APfVE}, p. 21.

\textsuperscript{50} \textit{APfVE}, p. 60.
also identify those who have been disadvantaged by forming their preferences in accordance with circumstances in which the opportunity for basic human flourishing was lacking. It is these preferences that count as adaptive.

The second advantage relates to identification and response. Without knowing which real-world preferences are adaptive, no ameliorative response can be undertaken. But if a list of components of basic flourishing can actually be developed, then persons' real preferences and circumstances can be compared against it in order to determine whether their preferences are adaptive. Khader's account will also require investigating the circumstances under which preferences were formed, but assigning content to human flourishing nevertheless simplifies the identification process immensely. Through a process of engagement or observation, identifying adaptive preferences becomes a matter of determining whether they meet some predetermined criteria, and armed with this information, development practitioners can direct their efforts in appropriate directions.

Nussbaum's account of adaptive preferences

Having noted these *prima facie* benefits of perfectionist accounts, I now turn to Nussbaum's account of adaptive preferences to analyze the drawbacks. Nussbaum does not herself see her account as a perfectionist one — indeed, as the previous references show, in her later work, she rejects perfectionism explicitly and completely. But her account of human capabilities, to which her account of adaptive preferences is tied, begins with a perfectionist account of the good and moves only gradually towards political liberalism, and her primary discussion of adaptive preferences occurs in the transitionary period in the late 1990s and early 2000s when her account seems to have both perfectionist and political liberal commitments. Nussbaum's account is particularly valuable here because of this explicit turn. Juxtaposing her political liberal justification of the capabilities approach with her perfectionist account of adaptive preferences can help to expose the drawbacks of accounts like the latter. Since her account is less explicit than Khader's, it cannot be as succinctly reviewed. But since it remains one of the most influential accounts of adaptive preferences in the literature, it is important to review it in some detail.

---

51 For similar statements made in relationship to her capabilities approach, see Martha Nussbaum (2006) *Frontiers of Justice: Disability, Nationality, Species Membership*, Cambridge, MA: Harvard University Press.
As Khader notes, discussions of adaptive preferences usually take the form of references to preferences that intuitively seem adaptive and rarely offer clear sets of conditions for determining adaptiveness. And although Nussbaum devotes the entirety of one of the four substantive chapters in *Women and Human Development: The Capabilities Approach* to the concept, she is no exception. In explaining adaptive preferences, Nussbaum offers a host of intuitive examples. Two of her favorites are Vasanti and Jayamma. Vasanti is an upper-caste Indian woman who remained for many years with an abusive husband who drank away their money and got a vasectomy without consulting her. Jayamma is a lower-caste Indian woman who performed physically demanding and poorly paid work in a brick kiln where the easier and better paid jobs were reserved for men. Nussbaum points to the two women’s lack of indignation and their compliance with their situations as examples of adaptive preferences. She points as well to women in Andhra Pradesh who lacked access to electricity, teachers, or bus services, but did not think that this was a bad thing, “that being the only way that they had known.” Similarly, in the area of health, Nussbaum points to a survey done after the Great Bengal Famine, in which widowers reported themselves to be in much worse health than widows in the same community. Since the two groups were independently judged to be at objectively similar levels of health, Nussbaum’s point is that the widows had come to expect much less for themselves and so to be satisfied with much less than their male counterparts. And although she does not point to empirical examples, she tells similar stories about those who do not know what it feels like to be well nourished and so are content with their undernourished state, and about women who lack a desire for education because of a lack of information, intimidation from their communities, or “adaptation to a view of life according to which boys are entitled to education and girls are not.”

52 APWE, p. 19.  
55 WHD, p. 140.  
56 WHD, p. 139.  
58 Martha Nussbaum (2006) *Frontiers of Justice*, Cambridge, MA: Harvard University Press, p. 279. Nussbaum makes many other similar references to adaptive preferences in her other works. For instance, see Martha Nussbaum (1995) “Human Capabilities, Female Human Beings” in Martha Nussbaum and Jonathan Glover, (eds.), *Women, Culture and Development: A Study of Human Capabilities*, Oxford: Clarendon Press, p. 91, where she points to both education and health as areas in which the “poor and deprived frequently adjust their expectations and aspirations to the low level of life they have known.” And in her most recent book on the capabilities approach, she offers as examples of adaptive preferences cases in which women do not form desires for education or work
Nussbaum’s account, however, is not exclusively intuitive. In a recent passage on the subject, she writes

People adjust their preferences to what they think they can achieve, and also to what their society tells them a suitable achievement is for someone like them. Women and other deprived people frequently exhibit such adaptive ‘preferences,’ formed under unjust background conditions. These preferences will typically validate the status quo.\(^5^9\)

This passage begins to offer a conceptual account. It specifies that adaptive preferences are adapted in order to fit in to available opportunities, that Nussbaum is concerned with preferences adapted to unjust conditions, and that adaptive preferences are problematic (at least in part) because they maintain an unjust status quo. But such an account also leaves many questions unanswered: are adaptive preferences always problematic? Are they problematic if they adapt to the status quo without endorsing it? Are preferences with the same content problematic if they are not adapted to circumstances? A complete account of adaptive preferences would require answers to these questions, and although Nussbaum’s earlier work provides clues, the answers suggested there do not readily cohere with the explicit political liberal position that she has adopted by the time she writes the passage above. Accordingly, I will work through Nussbaum’s earlier writings on the subject to try to recreate the fuller account that can most plausibly be attributed to her.

Before I do so, however, I should note that this rough, intuitive account of adaptive preferences likely suffices for Nussbaum’s own project. The purpose of Nussbaum’s capabilities approach is dual: it is meant to designate the appropriate space in which to make quality of life assessments of individuals, but it is also – and more importantly – meant to ground a set of entitlements owed to individuals, that she ultimately intends to be adopted as constitutional guarantees in all of the states of the world.\(^6^0\) Her project is quite ambitious: in addition to arguing that our focus should be the real opportunities that people have to do and be valuable things, Nussbaum wants to clarify the content of those opportunities, and assign duty-bearers who will be responsible for providing them. For her, it is governments that must guarantee their citizens these real opportunities, and it is the role of philosophers to aid them by


\(^5^9\) Nussbaum (2006) Frontiers of Justice, p. 73.

\(^6^0\) WHD, p. 12.
helping to devise an appropriately universal list that does not leave the content of the opportunities that individual persons happen to have up to chance.

Indeed, it is this insistence on providing a substantive universal list of central capabilities that is the most noteworthy feature of Nussbaum’s capabilities approach. From her earliest writings on the approach, she has insisted that furthering social justice requires providing a definite list of capabilities to which all individuals are entitled. And although Nussbaum’s route for arriving at the content on her list has been frequently criticized, her reasons for this commitment are important given the concerns about adaptive preferences that motivate this thesis. Because she wants the capabilities approach to “supply definite and useful guidance” to those involved in the work of fighting for social justice, and to “prove an ally in the pursuit of sex equality,” Nussbaum argues that it is not enough for states to concentrate on citizens’ capabilities, rather than their subjective well-being or shares of primary goods. Instead, she argues that social justice can only be achieved if citizens are guaranteed the right set of capabilities. Although she is happy to leave the provision of basic capabilities to individual states, she is not willing to allow the decision of which capabilities to provide to be made at the state level. Local deliberation might fail to recognize the central importance of some items on the list and lead to a failure to provide them. For Nussbaum, this trade-off is unacceptable. In her words, “some human matters are too important to be left to whim and caprice, or even to the dictates of cultural tradition.”

How, precisely, do adaptive preferences fit into this picture? Primarily, they motivate the project. It is important to focus on capability rather than actual desire because options that are inaccessible are often difficult to desire; it is important to have a universal list because those who already have adaptive preferences might fail to appreciate the value of the capabilities they lack and so might fail to demand their provision from their governments. To motivate the adoption of her universal list of

---


64 Ibid. p. 47.
capabilities, Nussbaum only needs to establish that there are some adaptive preferences that correspond to the items on her list. The examples that she employs do this effectively. And once her account is motivated, her attention is focused on giving a positive account of the contents of her list of capabilities, not on explaining in any precise terms what adaptive preferences are.

But while this primarily intuitive conceptual account of adaptive preferences may suffice to motivate Nussbaum’s own project, it is far too vague to provide guidance for either identification of or response to adaptive preferences. This is a problem because Nussbaum’s version of the capabilities approach has been adopted by development practitioners as a tool for addressing actual social injustice in the world.65 If they are going to use her work as a justification for responding to adaptive preferences, then they need a clearer version of her account so that they can more accurately judge whether it is in fact an account that recommends respectful treatment of persons. In what follows, I argue that it is not, on the grounds that it maintains too many of the perfectionist trappings that her broader account of human capabilities otherwise rightly rejects.

Perfectionist accounts of adaptive preferences: drawbacks

Although she prefers the term “Aristotelian essentialism”, Nussbaum’s early justification for the capabilities approach is clearly perfectionist. By “essentialism,” Nussbaum means “the view that human life has central defining features,”66 and her view is essentialist in that the capabilities on her list are meant to capture a set of functions that are intimately tied to a particularly human way of living. She is careful to make clear that her account of human capabilities should not be seen as an instrumental one (like, for instance, Rawls’s account of primary goods) that provides a set of resources that are to be valued for their ability to bring about valuable ends. Instead, the capabilities that Nussbaum discusses are understood to be ends in themselves – they are a constitutive part of the “overall shape and content of the


human form of life.” And not only are the capabilities ends in themselves – they are centrally important ends: they get at “what the most central features of our common humanity are, without which no individual can be counted (or counted any longer) as human.” So despite the fact that she does not use the term, her early work on the capabilities approach is clearly perfectionist.

But like both Khader and Rawls, Nussbaum recognizes that trying to find one complete conception of the good to endorse at the level of government would be both practically and morally undesirable. Practically, persons’ commitments regarding the good are simply too diverse for them to agree on one conception; morally, given Nussbaum’s emphasis on practical reason, it would not even be desirable to use the capabilities approach to enshrine one complete conception of the good, since doing so would impinge upon the fundamental human ability to plan a life and to act in accordance with that plan. So Nussbaum, like Khader, rejects the idea of a perfectionism concerned with excellences, and instead proposes a perfectionist theory of human good that focuses only on those things that constitute humanity at the most basic level. While we certainly disagree on what makes a human life excellent, she says, we do not disagree about what makes us human: here, she says, “we do share at least a very general outline of such a conception.”

How should adaptive preferences be understood in relationship to such a limited form of perfectionism? Since Nussbaum, unlike Khader, does not give a clear answer, we must guess. But we might suspect that her answer is similar to Khader’s in its reliance on its perfectionist content: adaptive preferences might be understood as preferences that fail to appropriately recognize the value of the items on the list. If the capabilities she lists form the core of the overall shape and content of the human form of life, then it makes sense that the adaptive preferences that those in charge of public policy should want to avoid honoring will be those that effectively prevent persons from living in these essentially human ways.

Defining adaptive preferences in this way, however, opens the concept up to the same objection that can be raised against the perfectionist account of truly human functioning that underlies it: both, in their perfectionism, might fail to respect persons. Although Nussbaum recognizes the importance of the human ability to plan a life and act in accordance with that plan, there are two ways in which this ability might be

67 Ibid., p. 215.
68 Ibid., p. 215, italics in original.
69 Ibid., p. 221.
70 Ibid., p. 215, italics added.
disregarded. Nussbaum uses the concept of adaptive preferences to explain why some persons will not endorse her list of central human capabilities. But a commitment to the truth of a list could license two different kinds of treatment of those who reject it. In the worst case, it could license paternalistic intervention aimed at coercing persons into the human functioning that they fail to prefer for themselves. This sort of paternalistic coercion clearly disregards the ability in question, because it denies persons the chance to act in accordance with their conceptions of the good. But even in the best case, where the failure to prefer the items on the list is not taken to license coercive intervention, defining basic human functioning still fails to recognize the first half of the ability – that is, to plan a life in the first place. By refusing to recognize as valuable those preferences that conflict with items on the list, it treats as deficient the plans of those whose preferences conflict.

At first glance, these two failures might seem to mirror the distinction made by Stephen Darwall between recognition respect and appraisal respect. The former consists in “giving appropriate consideration or recognition to some feature of [the object of respect] in deliberating about what to do”, while the latter consists in “an attitude of positive appraisal of [a] person either as a person or as engaged in some particular pursuit.” While the former is owed to persons by virtue of their personhood, the latter is discretionary. If a perfectionist list justifies paternalistically preventing persons from making decisions that do not accord with its content, then it fails to show them recognition respect, because it fails to show appropriate consideration for persons as moral agents for whom the ability to live a life in accordance with one’s own convictions has great value. Such recognition of persons as moral agents is indeed obligatory on any broadly liberal account of respect – at least absent countervailing considerations like the presence of threat of harm to others. For reasons given below, I call this kind of failure a failure of primary recognition respect.

Both Khader and Nussbaum recognize the disrespect that accompanies paternalistic coercive intervention, and both ensure that their perfectionist accounts avoid it. For Khader, this is as simple as recognizing that perfectionism can justify the rejection of coercive paternalistic intervention into individual lives on the grounds that coercion itself impedes basic human flourishing. This might be so either if personal freedom is itself taken to be central to flourishing, or if the frustration from having one’s preferences constantly frustrated outweighs the benefit of the flourishing.

---

72 *APWE*, p. 145. She offers other reasons as well, but this is the most wide-reaching.
behavior into which one is pushed. Either way, by prioritizing a particular conception of human flourishing, Khader’s account can show persons primary recognition respect.

Nussbaum’s solution is slightly different. As its name suggests, her capabilities approach makes the provision of capability rather than functioning the aim of public policy. Because the list is only meant to be one of opportunities that governments must provide and that individuals can choose whether or not to take advantage of, rather than a list of functionings that the government must impose on citizens, she holds that giving content to the list does not remove from citizens “the chance to make their own choices about the good life.” So the problem of primary recognition respect is solved on Nussbaum’s account because the emphasis on capability allows her to say that individuals who do not value one of the central capabilities will not be forced into acting against their values so long as the capability for the unvalued item remains firmly in hand.

So while neither perfectionist approach may approve of individuals non-flourishing-conducive choices, it still allows them to make those choices. And it might seem that this disapproval is a simple lack of appraisal respect, which is not similarly obligatory: even if showing recognition respect for a person requires me not to interfere with his decision to spend all of his time counting blades of grass, I am not under any obligation to show a positive attitude of appraisal for his choice to do so. Indeed, insofar as freedom of conscience and respect for persons are both important, it would be deeply problematic if showing respect for persons required each of us to approve of all of the choices that others made.

But while this is so at the individual level, the situation is different at the state level at which Nussbaum wants her list to be endorsed. When the state endorses a list of capabilities as a list of truly human functionings, it does not just disagree with those who would make the content of the list different; instead, it makes a statement at the societal level about the moral truth of the values that are used to structure society. And given the pervasive effects that the design of society has on individual lives, state


74 This is importantly not to say that the state makes a statement about moral truth every time that it commits a principle to law. For instance, when Article VI of the US Constitution states that the Constitution shall be the supreme law of the land, it does not thereby commit itself to the moral truth of the Constitution’s contents. The portion of the Declaration of Independence, however, that holds to be self-evident the truth that all men (sic) are created equal, does seem to commit itself to moral truth. When a list of capabilities is offered as an account of truly human functioning, it is more similar to the passage from the Declaration of Independence; it could also be offered as a list of entitlements owed to all citizens, and in this case it would be more similar to the statements made in the Constitution. I return to this distinction below.
endorsement is not as innocuous as individual endorsement. The problem with claiming truth is this: even a perfectionist state that shows persons primary recognition respect accuses those individuals who disagree with it of either moral laziness or moral ineptitude. What the state says to those who place no value in Nussbaum’s perfectionist conception is effectively this: “You do not appropriately appreciate what it is that makes your life a human one.” To be sure, we can all make mistakes about value, but when the state claims truth for its own conception, it allows content to trump the process of reasoning by which it is reached. No matter how thoughtful, informed, or considered a person’s judgment is, it will be taken to be woefully mistaken as long as it contradicts the content of a perfectionist list endorsed as true. In deciding how to evaluate the lives of citizens whose lives it constrains, the state that endorses a set of central human functionings fails to consider the interest that dissenting individuals have in being recognized as moral agents who can competently plan a life or arrive at a conception of the good. I call this a failure of secondary recognition respect.

A brief note on terminology: labeling these two types of respect “primary” and “secondary” might seem to suggest that one type is significantly more important than the other. To some degree, this seems correct. The ability to practice one’s religion, for instance, without threat or coercion is a deeply important good even in cases where others think that religion fundamentally wrong-headed. And indeed, we might also think that solving the problem of primary recognition respect is a prerequisite for solving the secondary problem, since it is hard to imagine how any recognition of a comprehensive doctrine’s legitimacy could be anything but token without the corresponding liberty to act in accordance with it. But although I call problems of the latter sort problems of “secondary” recognition respect, this should by no means signify that they are unimportant. In fact, in discussions of adaptive preferences, the secondary problem is at least as important as the primary one. When, for instance, Uma Narayan protests the tendency of Western feminists to treat women in traditional societies as “dupes of patriarchy,” she is protesting the implicit idea that

75 While Nussbaum is primarily concerned with the state, Khader is concerned with human development organizations. The influence of these organizations is not as extensive as the state’s, but insofar as they wield enough power to make changes in the communities where they work, the power dynamic between these organizations and communities is still highly skewed. Accordingly, what I say below about Nussbaum applies to Khader as well.

these women either cannot or do not form their own conceptions of the good. Similarly, Lisa Fuller criticizes accounts of adaptive preferences that "[suggest that] the oppressed cannot see how the injustices of the current order are reflected in the dominant attitudes and messages expressed within it." Fuller is protesting not constraints that are put on people's abilities to act as they see fit, but rather the either implicit or explicit suggestion that persons' moral capacities are not up to the task of differentiating between categories as important as the just and the unjust. And H.E. Baber has the same aim when she holds that we have reasons to see poor and deprived people as "rational choosers, coping as best they can and making the best of a raw deal" rather than as "irrational, subsisting in a 'slumbering state,' unaware of their rights, beaten down, victimized, and psychologically damaged." Clearly, Narayan, Fuller, and Baber's complaints would remain even in cases where persons were allowed to act on preferences deemed adaptive, and where primary recognition respect was therefore shown.

And although Khader and Nussbaum's perfectionist accounts both successfully show persons primary recognition respect, they cannot show the secondary kind. I begin with Khader, because she explicitly recognizes something like this objection. She claims that her perfectionist account neither indicates that individuals with adaptive preferences are morally deficient, nor that having a perfectionist conception of the good robs an individual of the opportunity to formulate her own conception of the good. But although these issues are relevant to secondary recognition respect, I will argue that they do not get to the heart of the problem.

First, Khader rejects the idea that "public endorsement of a conception of human flourishing ... require(s) treating persons with suspect preferences as though they were somehow morally deficient." She rejects this idea on the grounds that, contra numerous suggestions to the contrary, having an adaptive preference is first of all compatible with autonomy, and second of all compatible with having many other preferences that are nonadaptive. While I find the second (although not the first) of

79 APWE, p. 146-151.
80 APWE, p. 146.
81 The following authors either implicitly or explicitly hold that adaptive preferences are either incompatible with autonomy, or affect agents universally. Many of them hold both: Sophia R. Moreau (2004) "The wrongs of unequal treatment" University of Toronto Law Journal, 54(3), p. 304; Naila Kabeer (1999) "Resources, Agency, Achievements: Reflections on the Measurement of Women's Empowerment" Development and Change, 30(3); Anita Superson (2005) "Deformed Desires
these suggestions compelling, and will address both in detail in Chapters Four and Five, neither suggestion can solve the problem at hand. If we grant Khader’s first point, then an adaptive preference can be an autonomous one; it can never, however, be one that correctly identifies the good. And while autonomy is certainly one important good, it is not the only one: acknowledging the autonomy of a person’s preference will be little consolation when she is consistently told, no matter how deeply or in how many different circumstances she reflects, that her ideas about human flourishing are not and could never be anything but wrong. Similarly, recognizing that many of a person’s preferences are nonadaptive will be unlikely to satisfy the person who is especially deeply committed to the preference that is labeled adaptive. In short, acknowledging that a person often succeeds in recognizing human flourishing, and acknowledging that her failures are autonomous, still leaves us pointing to commitments that may be her deepest and most considered necessarily as failures. This remains an important failure of recognition respect.

Second, Khader rejects the idea that a perfectionist account of flourishing robs individuals of the chance to formulate their own conceptions of the good. Her response seems to be that perfectionism does not prevent individuals from forming their conceptions of the good because it still allows them plenty of space to engage with the good as an open question. In her words, “persons can have opportunities to publically discuss their values and be consulted about decisions that affect their lives without having to have personally participated in forming the most fundamental normative conceptions according to which their lives are judged.” This seems right, so far as it goes – if a perfectionist conception of the good establishes only a very limited set of goods, then persons are free to disagree about goods on which no stand is taken. But her response does not get to the heart of things – it can offer nothing to the person whose argument is precisely with the set of goods that has been put beyond the pale. And as long as we can imagine even one person, after critical reflection in many different situations, seeing one of these “goods” as no good at all in his own life, this counts as a failure of secondary recognition respect.


82 This point is separate from the question of whether or not an adaptive preference can represent a rational trade-off between either goods or bads.

83 *APWE*, p. 150-151.
While Khader's response to the problem of secondary recognition respect is to try to defend her account against it, Nussbaum's response is to abandon perfectionism altogether. In the late 1990s, Nussbaum rejects her previous perfectionist justification of the capabilities approach in favor of a political liberal one, for reasons that at times seem to relate to secondary recognition respect. In the next section, I will turn to this transition. Before I do, however, I offer two considerations to be taken away from this discussion of perfectionist accounts of the good. The first relates to secondary recognition respect:

**Consideration 3: Content neutrality.** An account of adaptive preferences should not be committed to a perfectionist, essentialist account of value, because such an account of value limits the ability of states and other actors to show secondary recognition respect for individuals who do not value the content given.

The other consideration concerns primary recognition respect and the extent to which perfectionist accounts might license coercion. As we have seen, Khader rejects coercive intervention entirely, on the grounds that coercion itself undermines flourishing. This claim about coercion, however, need not be shared by other perfectionist accounts. On Nussbaum's account, for instance, it is not clear whether or not coercive intervention will be licensed in the case of adaptive preference. To be sure, she makes it clear that persons must be allowed to refuse to exercise the capabilities on her list when they in fact possess those capabilities, but it is ambiguous whether this prohibition on coercion extends to cases in which capabilities themselves are not present. And indeed, other perfectionist accounts unapologetically license a very great degree of coercion. Séverine Deneulin, for instance, holds that "no development policy can be neutral with respect to a conception of the good," and celebrates Nussbaum's early perfectionism for recognizing this. She holds that the perfectionist conception of the good that she recommends should justify compelling people to live in accordance with the set of basic capabilities that are settled upon:

If development policies based on a perfectionist theory of the good seem paternalist, seem to restrict people's freedom to live the way they choose, so much the better, since that type of paternalism is nothing more than the refusal to see another person suffering from not being able to live a human life.85

85 Ibid., p. 516.
This willingness to license coercive intervention in the case of adaptive preferences is by no means necessarily attached to perfectionist accounts of adaptive preferences, but the variety of responses to the question should nevertheless make us sensitive to the problem. I offer a second consideration accordingly:

Consideration 4: Interference criteria. An account of adaptive preferences should not warrant undue interference into the lives of those with adaptive preferences; any justifiable interference must be consistent with showing individuals with adaptive preferences primary recognition respect, and an account of adaptive preferences should be developed in such a way as to ensure that it is compatible with showing this respect.

Transitioning from perfectionism to political liberalism

At this point, let us try to further refine Nussbaum's account of adaptive preferences. Clearly, Nussbaum thinks that adaptive preferences are a problem: they motivate her entire capabilities approach. But in our discussion of primary recognition respect, we saw that Nussbaum had no problem with persons who forgo one of the functionings on her list as long as they genuinely have the associated capability. This calls into question the earlier formulation of Nussbaum's account of adaptive preferences. Adaptive preferences do not seem to be any failure to appropriately value an item on the list. Instead, they seem to be preferences that both fail to show an appropriate appreciation of the value of the items on the list, and are also developed in situations in which the items on the list are not accessible. This would mirror Khader's second condition for the adaptiveness of a preference, namely that preferences be "formed under conditions nonconducive to [a person's] basic flourishing." In Nussbaum's parlance, this means that an adaptive preference would be a lack of a desire for an item on the list, formed absent the capability for that item.

But although these two accounts are structurally quite similar, Nussbaum and Khader justify their accounts very differently. While Khader offers her account of adaptive preferences only in relationship to a perfectionist account of the good, Nussbaum uses the idea of adaptive preferences both before and after her turn from perfectionism to political liberalism. Given this significant change in Nussbaum's position, we might expect her account of adaptive preferences to undergo a similarly significant change. I will argue however, that it does not. In the remainder of this chapter, my aim will be two-fold. My primary aim will be to draw considerations that will inform my own account from Nussbaum's turn towards political liberalism, but I

86 *APWE*, p. 51.
will do this by pursuing my secondary aim: to show that Nussbaum’s turn towards political liberalism seems at times incomplete, and that the account of adaptive preferences that seems to attach to her political-liberal position provides one of the strongest reasons to think that her account of the capabilities approach retains its limited perfectionism.

Let’s look, then, at Nussbaum’s turn towards political liberalism. In the late 1990s, she describes the turn as “a major change” in the way she understands the capabilities project:

I now understand the list of central human capabilities as a specifically political form of liberalism, in the Rawlsian sense. I imagine that citizens of many different comprehensive conceptions can all endorse the items on the list, as things that are essential to a flourishing human life, whatever else that life also pursues and values. It is neither an exhaustive account of the good nor a metaphysically grounded account [...]; it is not grounded in any theory of the human being that goes beneath politics.88

Like Rawls, she makes the turn on the basis of what Rawls calls the fact of reasonable pluralism: the fact that a plurality of reasonable but incompatible comprehensive doctrines is a permanent feature of societies with free institutions.89 Since people have a deep allegiance to their comprehensive doctrines, both Nussbaum and Rawls recognize that their respective projects are bound to fail if they require individuals to affirm some other comprehensive doctrine over their own. Instead, individuals must be able to affirm Rawls’s proposed theory of justice, and Nussbaum’s proposed list of capabilities, in a way that is compatible with continuing to affirm their own comprehensive doctrines.

But Nussbaum does not seem to make this turn as completely as Rawls. Even in this short passage describing the nature of her change, we find at least two ideas that stand in tension with one another: while she insists that her account is not metaphysically grounded and holds that her capabilities are not grounded in any

---

87 I say “seems to” because the highly intuitive nature of her treatment of the subject makes it difficult to say for sure what her account would be. It may well be the case that Nussbaum would in fact reject my interpretation. Even if she did, however, this would not pose a problem for my project. If I have filled in the contours of her account correctly, then the criticisms that I offer are legitimate. If I fill them in incorrectly, then I offer, at the least, an account that could be attributed to her with a high degree of plausibility. In that case, the indeterminacy of her account is problem enough. And in either case, my project is only secondarily exegetical – I analyze her account primarily in order to inform my own, and shortcomings that could plausibly be attributed to her account are to be avoided equally with those that are actual.


theory of the person that goes beneath politics, she also refers to the capabilities as "essentials of a flourishing human life." This tension is, I think, indicative of the fact that Nussbaum fails to appreciate two elements of Rawls's turn: the role of self-respect, and the way in which Rawls's account is political.

Self-respect plays a central role in Rawls's theory of justice. Indeed, he holds that principles of justice can only be acceptable if they secure for persons the social bases of self-respect, for unless they do so, he says, those principles "cannot effectively advance [persons" determinate conceptions of the good." This is because of the important function that he attaches to self-respect: it gives us, he says, "a firm conviction that our determinate conception of the good is worth carrying out." The importance of self-respect on Rawls's account has much in common with the idea of secondary recognition respect. Just as expressing secondary recognition respect requires recognizing persons as having the ability to competently form and act upon a conception of the good, a society that wants to ensure justice must facilitate the possibility that citizens can feel that their own determinate conceptions of the good are worth carrying out. And this connection between the two ideas should make it clear why perfectionism cannot express secondary recognition respect for citizens: perfectionism, having settled on a true account of human good, denies the truth of accounts that disagree — and without the possibility of seeing one's own determinate account of the good as true, it is very difficult to see it as worth carrying out. But of course, this shows only that Nussbaum's perfectionist account will have trouble expressing secondary recognition respect for citizens, not that her new turn towards political liberalism struggles to do so. To show this, I turn to the second element of Rawls's account that Nussbaum fails to appreciate: the way in which his liberalism is political.

Rawls modifies justice as fairness by making it into a decidedly political form of liberalism. This means that the liberal commitments that it makes apply only to a certain area of life: to the basic structure of a society, which includes its main political, social, and economic institutions. Since the scope of a political conception of justice is limited in this way, the commitments that it requires persons to make are similarly limited. In the political realm, political liberalism asks them to endorse a shared conception of justice, but in all other matters, it leaves their commitments to the

---

90 Ibid., p. 319.
91 Ibid., p. 318.
92 Ibid., p. 11-13.
discretion of their own comprehensive doctrines. Importantly, the way in which persons are required to see themselves is similarly limited: in outlining the way in which political liberalism conceives of the person, Rawls writes that “[w]hen we describe the way in which citizens regard themselves as free, we describe how citizens think of themselves in a democratic society when questions of political justice arise.” So Rawls’s political liberalism does require persons to see themselves in a certain way: it requires them to see themselves as citizens who, once again as citizens, share an interest in certain liberties, protections, and primary goods. How they see themselves in their broader role as persons, however, remains up to their own comprehensive doctrines.

It is not always clear that Nussbaum understands her capabilities approach to be political in quite the same way. In some places, her writing makes it clear that she explicitly intends her list to be taken in Rawlsian “political-liberal spirit… as a list that can be endorsed for political purposes, as the moral basis of central constitutional guarantees, by people who otherwise have very different views of what a complete good life for a human being would be.” Understood in this way, it does seem plausible that her list requires no controversial value commitments – especially since it is meant only to point to constitutional guarantees that governments should be bound to provide. And often, it seems that these are guarantees that are owed, as Rawls’s constitutionally enshrined two principles of justice would be, to persons in their capacity as citizens.

But in other places, Nussbaum’s work is much less straightforwardly Rawlsian. Take the following passage: “[T]he central capabilities are not just instrumental to further pursuits: they are held to have value in themselves, in making the life that includes them fully human.” Or, a passage written at a similar time, in which she states that hers is a “specifically political theory that is only partially comprehensive.” Here, it is much more difficult to understand the capabilities as those things that citizens must be required to value for themselves and their fellows in the political realm. Instead, Nussbaum seems to be asking persons to affirm something much more controversial – she asks them to take a stand on what it is that, at least at the most

93 Ibid., p. 33, emphasis added.
94 WHD, p. 74.
95 Nussbaum often speaks in this way in WHD. For instance: “the capabilities in question are important for each and every citizen” p. 6; “a list of human capabilities that are indispensable for any citizen” p. 165; “What would it be to treat each citizen as an end, promoting the full range of the human capabilities?” p. 252.
96 WHD, p. 74.
basic level, makes their lives human. And by taking a stand on a point like this, we become unable to provide the social bases of self-respect and so unable to show secondary recognition respect for persons: in taking a stand on what makes lives human, we make it much more difficult to see as true—and so worth carrying out—those conceptions of the good that disagree about what makes lives human.

If there is ambiguity as to which of these two options Nussbaum intends—that capabilities be seen as things owed politically to citizens, or as things that persons are meant to see as the most basic elements of a good human life—her discussion of the structure of capability supports the latter. Once again, Nussbaum allows persons to forgo functioning as long as they possess meaningful capability. But as I noted, it is ambiguous whether they are similarly allowed to refuse functioning when they do not possess the capability in question. Accordingly, much may hang on what is involved in actually having a capability.

Nussbaum states in no uncertain terms that the capabilities on the list must not be understood in a “purely formal manner.” A person who formally has the right to an education does not have the capability to actually be educated if she lives so far from the nearest school that she cannot reach it, or if she cannot afford the required uniform, or if the teachers in her school rarely show up to teach. Instead, meaningfully having a capability requires having access to the material conditions that make the associated functioning a real possibility.

But capability requires more than external resources. On her account, robust capability has three parts: basic capability, internal capability, and combined capability. A basic capability is the latent internal capacity that can be developed into a capability. For instance, nearly all human beings are born with the capability to develop language, and therefore literacy. But basic capability is not full capability—if we give a child with only this latent capability a book, she cannot tell us what it says. So capabilities must also have internal components. Nussbaum calls these “developed states of the person herself that are, so far as the person herself is concerned, sufficient conditions for the exercise of requisite functionings.” Yet not even internal capability is sufficient. Full capability is combined capability: that is, internal capability in conjunction with the external resources to exercise the associated functioning.

---

98 For examples of these statements, see WHD, p. 41, 92.
99 WHD, p. 86.
100 WHD, p. 84.
101 WHD, p. 84.
person who is literate may have the ability to read, but she does not have the capability
to do so unless she has access to books or other print materials.

Nussbaum herself uses the example of literacy to introduce this tripartite
 distinction, and its qualities make it an extremely useful tool for explicating the
 concepts involved. Jointly, they give us a good sense of the point at which a person has
 a capability, and accordingly, the point at which she can permissibly choose not to
 exercise it: in order to meaningfully have the capability to read, a person must at
 minimum be literate, have access to books, and have the time to read them. Without
 any one of these components, the “choice” not to exercise her capability is, in an
 important sense, a false one.

But I worry that the example of reading, precisely because of its neatness, might
 obscure in other less neat cases what it is that having a capability requires. For
 instance, what does it mean to have the capability to show concern for other human
 beings? It seems plausible that, evolutionarily, human beings have developed the
 basic capability to care for others, and it seems correct that certain external or material
 conditions are required in order to meaningfully show concern for other human beings
 (including, at minimum, being in a position to interact and communicate with others).
 But what might internal capability amount to in this case? Imagine for a moment that
 I have the basic capability to care for others, and that I am, as well, surrounded by
 other human beings, so that I have the external requirements for the exercise of the
 associated functioning. But imagine further that I have been raised as a full-throttled
 ethical egoist, whose parents met my basic material needs, but showed me little love
 and taught me to think exclusively of my own interests. If I take on this value set
 wholeheartedly, then it seems to me implausible to say that my inborn capacity for
caring and my proximity to others really gives me the capability to care for other
 human beings. In failing to recognize the value of caring, I am, in a meaningful sense,
 unable to care.

Many of the capabilities on Nussbaum’s list have similar valuational
 requirements in their internal components: it is hard to understand how one can be
 able to be treated as a dignified being whose worth is equal to that of others if she does
 not see herself as the equal of those around her, difficult to see what it means to have the
 capability for sexual satisfaction if one sees sexual satisfaction as a fundamentally
 dirty or despicable thing, and hard to understand how one can actually live with

---

102 This is part of the capability for affiliation, number 7 on Nussbaum’s list. WHD, p. 79.
103 Or as Nussbaum often says, “pleasure.”
concern for and in relation to the world of nature if she does not see nature as a thing of value.\textsuperscript{104} Although Nussbaum does not herself make this point about value in her discussion of internal capabilities, statements that she makes in the rest of the book support it. Early in the book, she cautions that life circumstances “affect the inner lives of people, not just their external options: what they hope for, what they love, what they fear, as well as what they are able to do.”\textsuperscript{105} So it seems right that internal capability should, for Nussbaum, include not just technical skills like the ability to recognize letters and string them together into words, but also the space to value the ability to see an option as something to be hoped for, something to be loved, something the loss of which is to be feared.

This point is at the heart of the concept of adaptive preference. It is not only lack of external resources that keeps persons in inadequate circumstances; it is also the inability to see distant opportunities as ones that a person could actually consider for herself. This will matter for a conceptual account of adaptive preferences as much as it will matter to those engaged in the practice of response: the inability to recognize meaningful alternatives in the world will both raise the level of suspicion that an adaptive preference deserves, and be an appropriate locus for responses aimed at ameliorating adaptive preferences. Accordingly, I offer a third consideration to inform my own account:

\textbf{Consideration 5: Live options.} In determining whether the absence of a particular behavior reflects an adaptive preference, an account of adaptive preference should take into account the extent to which a person sees that behavior as a live option for herself.

But despite the importance of this point, thinking about capability in this way brings Nussbaum back towards a perfectionist account of basic human good. If having a capability is meant to be a prerequisite for meaningfully deciding not to exercise a functioning, and if actually having a capability requires appreciating the value of the thing in question, then in order to meaningfully decide not to engage in a functioning on the list, we seem to end up having to say once again that an individual must endorse the value of all of the goods on the list.

And indeed, if we look to Nussbaum’s discussion of adaptive preferences, it is here perhaps most clearly that we find reason to see her account as maintaining its limited perfectionist commitments. Earlier, I offered this formulation of Nussbaum’s

\textsuperscript{104} All of these are capabilities on Nussbaum’s list. For these and the others, see \textit{WHD}, p. 78-80.

\textsuperscript{105} \textit{WHD}, p. 31.

38
account of adaptive preferences: an adaptive preference was a lack of a desire for an item on the list, formed absent the capability for that item. At this point we are better positioned to see what such an absence of capability amounts to. I propose that we now see adaptive preferences on Nussbaum’s account as failures of internal capability that prevent people from appropriately valuing the substance of the capabilities on her list.

Such an account is borne out by much of what Nussbaum says on the subject. In preface to her discussion of adaptive preferences, Nussbaum notes that

Where there is lifelong deprivation, the distinction [between internal and combined capabilities] is not so easy to draw: persistent deprivation affects the internal readiness to function. A child raised in an environment without freedom of speech or religion does not develop the same political and religious capabilities as a child who is raised in a nation that protects these liberties.¹⁰⁶

This passage suggests that it is indeed internal capability that is at stake in instances where adaptive preferences exist. And later in her discussion, Nussbaum recommends that in cases where people do not exercise capabilities that they seem to possess, we should refrain from actually attributing capabilities to them until we have determined whether “persistent inequalities or hierarchies may have created emotional barriers to full participation.”¹⁰⁷ Although she does not use the term “adaptive preference” here, the implication is that persons who are not emotionally able to exercise a capability lack that capability as surely as those who lack the external resources to do so. And although there may be several kinds of emotional barriers to the exercise of a capability, a failure to appropriately value the capability seems to capture what Nussbaum has in mind when she speaks of Vasanti failing to protest her husband beating her, or Jayama quietly accepting the lower-paid work in the brick kiln, or girls thinking that education is owed only to their brothers. In each of these cases, there may or may not be a complete failure to value the capability in question – Vasanti, for instance, may simply be unable to imagine a life without the abuse, rather than thinking that there is nothing wrong with the status quo. But even if her failure is only one of imagination, it still seems to prevent her from appropriately valuing the corresponding capability to be free from assault.

Calling these sorts of preferences adaptive may not seem to raise the problem of secondary recognition respect, since it seems likely that the individuals in each case would not be deeply committed to their preferences. But what if they were? What if we

¹⁰⁶*WHID*, p. 85.
¹⁰⁷*WHID*, p. 93, emphasis added.
are trying to decide whether or not the preference of a woman to have her genitals cut is adaptive? It is not difficult to imagine that the woman in question genuinely finds uncut female genitalia profoundly ugly.\textsuperscript{108} It is here, when individuals are \textit{committed} to values that conflict with those on the list, that the problem of secondary recognition respect really arises. Even if a perfectionist who opposes female genital cutting\textsuperscript{109} holds that women should not be coercively prevented from undergoing surgery, the requirements of secondary recognition respect are violated when those women are told that their firm convictions are not and could never be compatible with their flourishing. Even if they can act on their convictions, the social bases of self-respect will be closed to these women. Yet this seems to be precisely the commitment that Nussbaum makes from within her political liberal position. In one of her clearest passage on the subject, Nussbaum states that adaptive preferences should be defined by their substance. Her version of the capabilities approach, she says,

addresses the problem of adaptive preference, again, by substantive rather than formal devices, as seems necessary. A habituated preference not to have one of the items on the list (political liberties, literacy, equal rights, or whatever) will not count in the social choice function, and an equally habituated preference to have such things will count.\textsuperscript{110}

To be sure, Nussbaum’s account of the capabilities approach in the late 1990s and early 2000s is ambiguous – at times it seems as if her commitment is to a limited perfectionism, while at others it seems that her commitment is indeed politically liberal. As I noted at the beginning of this chapter, this ambiguity has largely disappeared in her more recent work, and given the importance of secondary recognition respect, I think that revision away from perfectionism represents a move in the right direction. But her account of adaptive preferences does not undergo a similar revision. Although Nussbaum continues to discuss the problem of adaptive preferences in her later work, she does not provide a clear way of thinking about them that avoids the perfectionist leanings that I have described.


\textsuperscript{109} As both Nussbaum and Khader do. See \textit{WHID} p. 87, 94; \textit{APWE} p. 187.

\textsuperscript{110} \textit{WHID}, p. 149. This commitment to the substance of the list is enforced when she calls preferences that conflict with her list “corrupt and mistaken” (\textit{WHID}, p. 160), and “inadequate” and “unwise” (\textit{WHID}, p. 161).
Conclusion

In this chapter, I have looked at two perfectionist accounts of adaptive preferences in order to begin to develop the list of considerations that will inform my own account. Before I end the chapter, I will add one more.

We have seen that Nussbaum’s account of adaptive preferences, like Khader’s, is best thought of in relationship to a limited but substantive perfectionist account of human good. Such accounts still have the advantages that I noted at the beginning of the chapter: they provide an appropriate conceptual reason for treating adaptive preferences as problematic, and they provide substantive guidelines for identifying adaptive preferences in the real world so that they might be responded to. But we can now see that these advantages come at too high a price.

The price is too high for two reasons. The first has already been well-documented: by defining adaptive preferences in relationship to a substantive perfectionist account of the good, we commit ourselves to that associated conception of the good, and as I have argued, such a conception of the good cannot extend full recognition respect to persons. Accordingly, Consideration 3 already cautions against adopting an account of adaptive preferences that commits us to a perfectionist account of the good.

But there is a second reason, as well, that perfectionism is too costly: by defining adaptive preferences in relation to a substantive perfectionist account of the good, we hold the content of preferences hostage to adaptiveness—in other words, we make it impossible for a preference with a certain content to ever be nonadaptive. Of the two perfectionist accounts that we have looked at, Nussbaum’s account does this most explicitly. If I am right that her account of adaptive preferences is best understood as a failure of internal capability that prevents a person from appropriately valuing the substance of the capabilities on Nussbaum’s list, then a person can only develop a nonadaptive preference if she remedies this failure of internal capability by coming to appreciate the value of the capability in question. But once she has done this, her values will no longer conflict with the values on the list—she will be able to shed her adaptive preference, but only by changing it.

For Khader, the situation is slightly more complicated. As we saw, while the first condition of Khader’s account of adaptive preferences is substantive, the second is formal: to have an adaptive preference, a person must fail to prefer her own flourishing, but must also have developed that preference under conditions that are nonconducive to her flourishing. This may make it seem as if adaptive preferences are
not hostage to their content on her account, and indeed, Khader offers the example of a person with a preference for bullfighting to make it clear that preferences that are nonconducive to flourishing should not be counted as adaptive on that count alone. But while the two conditions can come apart in cases where they have never come together, the reverse is not true. Once they have come together, they cannot come apart: while Khader’s third condition for the adaptiveness of a preference is an expectation that it would change under conditions conducive to flourishing, she recognizes cases where this change does not occur as instances of particularly persistent adaptation, rather than the transformation of an adaptive preference into a nonadaptive one. So on Khader’s account as well, adaptive preferences remain hostage to their content.

We use the idea of adaptive preferences because we value the freedom of individuals to live according to the values that matter to them. We consider the possibility that some preferences might be adaptive because we think that there are certain sets of circumstances that can prevent people from valuing the sorts of things that they would be likely to value in nearly any other set of circumstances. But if persons do continue to value the same things after much thought and in these various situations, then taking seriously the freedom of individuals to live according to their own set of values should require us to see these preferences as no longer adaptive. And when we define adaptive preferences by reference to their content, we are unable to do this. An account of adaptive preferences should alert us to situations in which persons may not have the space to interrogate their preferences, in order that such space might be made; it should not encourage us to categorize preferences as very persistently adaptive in situations where this space has come into existence and preferences have been interrogated and maintained. So at this point, I propose one final consideration:

**Consideration 6: Transformation of preference.** In order to respect individual freedom, an account of adaptive preferences should allow for the possibility that some preferences previously counted as adaptive may become nonadaptive without any change in their content.

---

111 *APWE*, p. 48.
112 For instance, she notes that “some people probably have internalized [adaptive] preferences so deeply that there is no hope of changing them” (*APWE*, p. 19), and that we “cannot reveal with absolute certainty whether people’s preferences will change under better conditions; deprived people may find that they are more deeply attached to self-deprivation than even they knew” (*APWE*, p. 185).
Chapter Two: Deliberative Accounts of Adaptive Preference

In this chapter, I develop a set of considerations from the work of deliberative theorists who are concerned with the concept of adaptive preferences. Although adaptive preferences play a primarily motivating role in their work, much as they do in Nussbaum’s, the proposals of these deliberative theorists suggest a promising route for identifying and conceptualizing adaptive preferences in a way that shows persons secondary recognition respect. This ability, however, comes at a cost: it is unclear that such accounts can expose adaptive preferences to the kinds of scrutiny that would be required to uproot them.

Deliberative projects

In this chapter, I once again focus on two theorists whose work is especially well-suited to inform a discussion of adaptive preferences: Amartya Sen and Brooke Ackerly. As I said in the Introduction, a full discussion of adaptive preferences will have three parts: a conceptual account of what adaptive preferences are, an account of how the preferences identified in theory by the conceptual account should be identified in practice in the real world, and an account of how the preferences identified in the real world should ultimately be responded to. Like Nussbaum, these deliberative theorists use the concept of adaptive preferences primarily in order to motivate their further projects of identifying and responding to preferences that are more broadly problematic. Accordingly, what can be taken away from this discussion of deliberative accounts will be most relevant in the areas of identification and response, and will contribute less to the development of a conceptual account of adaptive preferences. As we will see in Chapter Five, however, discussion of the potential failures of deliberative theorists to respond to more broadly problematic preferences will also help to inform the conceptual account of adaptive preferences that I ultimately develop.

These deliberative theorists are concerned primarily with how members of a group can identify a set of robust values to guide their action and how they can decide on the basis of those values what course of action to take. Adaptive preferences enter the equation because both theorists write in development contexts where they recognize that limited and unjust circumstances can cause persons to adopt values and
practices that they would not otherwise endorse and that can therefore serve as problematic reasons for action. Sen has often pointed to this problem, repeatedly making observations like the following:

"Desires reflect compromises with reality, and reality is harsher to some than to others. The hopeless destitute desiring merely to survive, the landless laborer concentrating his efforts on securing the next meal, the round-the-clock domestic servant seeking a few hours of respite, the subjugated housewife struggling for a little individuality, may all have learned to keep their desires in line with their respective predicaments."

Given that people make such compromises, he argues, it would be "ethically deeply mistaken" to take such preferences at face value, or to allow them to determine the ways in which persons ought to be treated. Ackerly similarly notes that persons who are too close to practices and norms that support harmful inequalities "may not recognize them as harmful or be able to be critical of them."

However, both theorists also oppose the perfectionist's proposal to judge persons' preferences and behaviors against some external, supposedly objective criteria. Instead, they promote collective deliberation aimed at improving and clarifying persons' beliefs and values by exposing them to reasoned scrutiny, so that persons can be empowered to direct the course of their own lives. Although Ackerly argues that social criticism should be guided by a substantive list much like Nussbaum's, which seeks to capture the kinds of choices that all persons ought to be able to make in order to live lives worthy of human beings, she holds that such a list should inspire deliberation rather than constrain its outcomes. Ackerly calls such deliberation doubly valuable, since it both promotes intrinsically valuable learning and understanding, and also provides previously silenced persons with the tools to "promote deliberative inquiry in the larger society where their views have previously been excluded."

Sen also argues extensively for the importance of freedom and self-determination in human life, and supports deliberation as a way of improving persons' values, writing that any plausibility that value claims might have is


"dependent on their ability to survive and flourish when they encounter unobstructed discussion and scrutiny."[119] If beliefs and preferences cannot survive such scrutiny, then individuals will be confronted with the fact that they lack reason to hold them, and they should accordingly become less broadly adopted. David Crocker, expanding upon Sen's deliberative theory, praises it for its ability to "both [take] seriously people's judgments and [subject] them to collective rational scrutiny."[120]

Clearly, this shared deliberative project is not directed solely at adaptive preferences. While the adaptation of preferences to unjust or inadequate circumstances represents one way in which preferences might become corrupted and so require deliberative scrutiny, many other kinds of preferences may also require such scrutiny. Foolish, irrational, or lazily formed preferences would all be as likely to be rejected in this process of public deliberation as would adaptive preferences. Nevertheless, since adaptive preferences fit into the set of preferences to which these deliberative projects are meant to respond, examining those projects will still provide insights into appropriate and effective ways to respond to the more limited set of adaptive preferences.

Benefits of deliberative response: efficacy, agency, autonomy

As I noted in the Introduction, it is somewhat artificial to separate the three elements of concept, identification, and response. This will especially be the case when it comes to deliberative conceptions of adaptive preferences. Persons can certainly deliberate about values in the abstract, but the proposals advanced by Ackerly and Sen are aimed at evaluating the way in which real persons live their lives, and are therefore focused on specific, concrete questions about value. Since such questions are hard to consider without a corresponding idea of the alternatives available, a discussion of which preferences ought to count as adaptive (or more broadly problematic) will already bleed into a discussion of how potentially adaptive (or otherwise problematic) preferences ought to be replaced.

A deliberative approach to response, however, need not require a deliberative approach to identification. As we saw in the last chapter, Khader's procedure for identification of adaptive preferences is primarily perfectionist. It relies on a


predetermined substantive list to distinguish adaptive preferences from nonadaptive ones. Nevertheless, she explicitly recommends a deliberative approach to response to adaptive preferences when she offers only a minimal and vague account of human flourishing. Since it takes no stand on how the basic components of flourishing can concretely be realized, the kind of account that she proposes "underdetermines what strategies for adaptive preference change should be chosen in any case."\textsuperscript{121} In other words, while a vague and minimal account of the good can provide the criteria for determining whether a preference is flourishing incompatible and therefore a candidate for adaptiveness, it cannot rank the many flourishing-compatible possibilities that might replace that preference. Instead, Khader argues that community members who previously held the preference in question must themselves deliberate to determine which possible alternative is most desirable given their own particular history, values, and circumstances. So at the response stage, her theory is genuinely deliberative.

Khader, however, would also call her own theory deliberative at the identification stage, for two reasons. First, she argues that the substantive content of the list that she proposes should be developed as part of a deliberative, global, cross-cultural exercise.\textsuperscript{122} And second, she recommends that development practitioners engage in deliberation with community members in order to determine whether their suspected adaptive preferences \textit{are} in fact adaptive.\textsuperscript{123} Yet I am hesitant to call her identification strategy deliberative: while it certainly \textit{involves} deliberation with affected parties, it imports an account of value that does not seem open to question by affected parties at the local level. While it is open to community members in the second round of deliberation to argue that their suspected adaptive preferences have been misinterpreted and do not in fact violate the content of the list given, it does not seem to be open to them to question whether the content of the list itself has been mistakenly selected. I say that the content "does not \textit{seem} to be open to question" because Khader does not explicitly address this issue. While she could plausibly side with deliberative theorists like Brooke Ackerly,\textsuperscript{124} who incorporate substantive lists into deliberation but make it clear that the content of such lists should itself be the focus of discussion at the community level,\textsuperscript{125} Khader's commitment to a perfectionist

\textsuperscript{121} Serene Khader (2011) \textit{Adaptive Preferences and Women's Empowerment} Oxford: Oxford University Press, p. 21.
\textsuperscript{122} Ibid., p. 20
\textsuperscript{123} Ibid., esp. p. 63 and following.
\textsuperscript{125} "Community level" is, admittedly, a rather ambiguous phrase. When I use it, I mean it to refer to deliberation undertaken by those most directly affected by the preference or decision in question. I
account suggests that she sides more closely with a theorist like Nussbaum who stipulates that a habituated preference not to have one of the capabilities that she proposes will not count in the social choice function. Otherwise, her commitment to perfectionism seems to add little to the discussion in question. For the moment, then, I call only Khader’s response account deliberative, and her work will therefore play a comparatively small role in this chapter.

Deliberative response accounts do, however, have several important benefits worth noting. First, state and non-state actors that respond to adaptive preferences should be concerned with the efficacy of their programs. Deliberative responses are likely to be comparatively efficacious, since adaptive preference interventions designed with little or no input from community members may waste both time and money. This waste, when it occurs, may be due simply to a lack of expertise. Individuals have knowledge of their own circumstances that can be difficult – if not impossible – to match from the outside, and interventions that are designed in the absence of deliberation with affected parties may be less effectively designed than those that take account of local insight and analysis. If the needs of affected individuals are not known then they are simply unlikely to be met, and deliberating with affected individuals about their needs is probably the single best way to bring previously unrecognized needs to light. Otherwise, persons may simply ignore the interventions proposed.

More extremely, interventions that are mounted without the input of community members may even be counter-productive if those affected disapprove of the intervention or become determined to resist it. For instance, even some Afghani women who resent the Taliban’s imposition of the burqa also resent Western pressure to remove it. In an interview, one woman said that “[w]e are not going to take it down because the day before we didn’t wear it.” Unless community members

recognize, however, that this set of affected people may be very different depending upon the preference or decision at issue, and might range from a few affected persons in a local neighborhood to all of those persons living within the confines of a state, or even, at the limit, all living and future human beings.

128 Women who themselves opposed female genital cutting in Senegal urged the government not to outlaw the practice, arguing that women who did not endorse the law would simply ignore it. Asma Mohamed Abdel Halim “Fram Bagadadji to Abu Hashim: New Approaches to Combat Female Circumcision” Review of African Political Economy, 24(114).
take interventions to be valuable, they are unlikely to comply with them, and as in the case of these Afghaniwomen, may even actively resent or resist them. Interventions aimed at changing practices supported by adaptive preferences, then, might be ineffective or do more harm than good if the communities targeted do not endorse the intervention.

The benefits of deliberating with persons with adaptive preferences, however, go beyond the success or failure of isolated projects. Deliberating increases the agency of those involved, especially when that deliberation leads to and informs action that is subsequently taken. Influencing the actions that affect their own lives allows persons to take control of those lives in a way that is deeply important for those who have for a long time experienced powerlessness. Indeed, in the participatory poverty assessments undertaken as part of the World Bank's *Voices of the Poor* project, poor people consistently ranked the sense of powerlessness and dependency as one of the most pernicious aspects of poverty. The sense of efficacy that comes from successfully designing and launching a project that affects one's own life can pervade the rest of that person's life, and have effects far beyond the project at hand.

But perhaps most importantly, the very process of deliberating together with others can enhance one's capacity for autonomy. I will give a fuller account of the concept of autonomy in Chapters Four and Five, but for now it is enough to note that virtually all accounts of autonomy share an emphasis on the capacity to critically reflect upon one's own desires, values, and priorities. Engaging with others to decide on new values and preferences to replace an old adaptive preference will necessarily require individuals to call on one another to explain and give reasons for their positions, and the strengthened capacity for this sort of careful and critical consideration should extend beyond the situation at hand to the other areas of individuals' lives. Ideally this enhanced capacity for critical thought will encourage persons to reflect independently not only on the preferences in question, but on their other preferences as well. In this way, deliberation can be beneficial even when it is not acted on, or when actions taken disregard the outcome of deliberation.

Given these benefits of deliberative response accounts of adaptive preferences, I propose two further considerations to inform my own account of adaptive preferences:

---

130 For instance, the Indian practice of *sati* was relatively rare until the British tried to eradicate it. At that point, it became "a larger-than-life symbol of 'Hindu' and 'Indian' culture" for both Indians and British." Uma Narayan (1997) *Dislocating Cultures: Identities, Traditions, and Third World Feminism*, New York: Routledge, p. 65.

Consideration 7: Efficacy. A theory of adaptive preferences should be conducive to providing solutions that those affected will themselves endorse, since only solutions that affected persons endorse have a significant likelihood of being effective.

Consideration 8: Facilitation of Relevant Skills. Where possible, a theory of adaptive preferences should recommend responses to adaptive preferences that strengthen the capacities of agents beyond the situation at hand.

Benefits of broadly deliberative accounts: provision of secondary recognition respect and maintenance of standards for moral reasoning

Admittedly, since Sen and Ackerly treat adaptive preferences as a subset of the problematic preferences that might be identified through deliberation, they do not provide explicit conceptual or identification accounts of adaptive preferences. Identification of broadly problematic preferences, however, looks something like this: using procedural rather than substantive means to interrogate the preferences of actual individuals and communities, deliberative theories determine in particular cases whether preferences are problematic by discovering whether they will be maintained in the face of reasoned questioning. As we saw in the Introduction, adaptive preferences must be adapted to circumstances which are in some way problematic, and deliberative attempts to identify adaptive preferences specifically could leave it up to the deliberative process to determine which previous circumstances counted as problematic and which current preferences seem to have been developed specifically in reaction to them. This suggests something like the following conceptual account: An adaptive preference is that one that has received insufficient or uninformed critical reflection as a result of unacceptable circumstances, and that would be rejected if this reflection took place. Proponents of this kind of account would all agree that socialization can cause us to prefer things that are not good for us, but nevertheless maintain that the ultimate judgment of whether any particular preference is harmful must be made by the person who holds it.

This kind of deliberative account has two additional benefits. First, such accounts avoid the major problem faced by perfectionist theories of adaptive preferences. Recall that perfectionist theories, given their commitment to a substantive account of human flourishing, had difficulty showing secondary recognition respect to persons with conceptions of the good that conflicted with some of the elements of the account proposed. This trouble arose because perfectionist accounts begged the question against conceptions of the good that deny the value of items on the list or
endorse items that conflict with that list, by counting them as necessarily adaptive. Deliberative identification accounts, on the other hand, beginning as they do without a substantive account of value, do not face the same problem. Instead, they rely on the ability of values, norms, and practices to stand up to critical scrutiny, thereby at least theoretically allowing that preferences arising from conceptions of the good with all possible contents could be nonadaptive. Deliberative identification accounts, then, seem at least to be superior to perfectionist identification accounts in their ability to show persons secondary recognition respect.

We should, however, revisit for a moment exactly what secondary recognition respect requires. In the first chapter, I held that secondary recognition respect involved giving appropriate consideration or recognition to the interest that individuals have in being recognized as moral agents who can competently arrive at a conception of the good. But while individuals have this interest, showing consideration for it does not, I think, require ignoring evidence about persons' actual moral reasoning abilities. Some persons do not conduct their moral reasoning in a competent manner: they are unable to reason critically, or they do so lazily, or in willful ignorance of some important relevant information. Indeed, virtually all of us will, at times, reason in a suboptimal manner. If we suggest that this is not the case, and that moral reasoning cannot be done in better and worse ways, then we are thrown back to a position of moral subjectivism in which persons with differing conceptions of the good cannot have reasoned moral conversations at all. We should not, then, deny the possibility of poor moral reasoning, or evidence that suggests incidences of it, in an effort to respect persons. Instead, we should be careful to ensure that we are using the right kind of evidence to determine whether a person's interest in being recognized as a competent moral reasoner is being appropriately overridden. The problem with perfectionist accounts is not that they classify some persons as poor moral reasoners, but that they apply this classification automatically to persons who might be thoughtful, informed, and reflective about their conception of the good and accompanying preferences. In doing this, perfectionist accounts ignore the disagreement that characterizes discussion of the good in any plural society and effectively claim unwarranted privileged epistemic access to the answers to those questions. Deliberative identification accounts, on the other hand, also allow for the possibility that individuals may at times conduct their moral reasoning poorly or

---

132 At least insofar as they have been developed in accordance with problematic social circumstances.

50
lazily, but do not similarly dismiss the moral disagreement that comes with living in a plural world. Any suggestion of an individual’s poor moral reasoning will be based not on the content of a person's conception of the good, but on her unwillingness or inability to recognize reasons against her own account, or to offer reasons in its defense. This is the second benefit of deliberative identification accounts: they allow us to maintain standards for collective moral reasoning without ruling any person’s conception of the good out from the start on the basis of its content.

**Two theoretical weaknesses: inability to respect faith-based moral reasoning, and indeterminacy of the point of success**

However, a problem related to secondary recognition respect may still remain. I have already argued that it is unacceptable to place substantive constraints on the content of preferences that persons can hold non-adaptively, but one could hold, further, that there is even something wrong with insisting that persons go through a certain procedure in order for their preferences to be counted as non-adaptive. Thomas Pogge, for instance, argues that “I respect someone’s autonomy only insofar as I accept his measure of flourishing as well as his way of arriving at this measure – without demanding that he must have come to it on some path that I approve as sufficiently reflective.” What Pogge seems to be suggesting is that we fail to respect persons when we place external requirements on the way in which they arrive at the preferences that we count as meaningfully their own.

This argument has the most bite, I think, when it is raised by those who hold that critical questioning can detract from rather than contribute to the possession of some valuable knowledge. This includes those religious believers who think that it is precisely the hallmark of religious faith to believe in that which cannot be proved – and may even be brought into question – by reasoned argument. By raising this issue in the context of religion, I do not mean to suggest that the “correct” interpretation of any religion requires this unwillingness to question its tenets, or that persons only

---

Thomas Pogge (2008) *World Poverty and Human Rights* (Second edition), Cambridge: Polity Press, p. 36-37, italics in the original. Note that this question (namely, whether we should require some reflective procedure before recognizing a person’s measure of her own good as authoritative) is different from the question of whether we can require some sort of collective deliberative procedure (rather than, say, a simple voting procedure) before a collective action is undertaken. Pogge’s commitment on the first point does not entail any commitment on the second point.

For an academic account of such a position, see William James, who argues that evidence for religious belief may only become available to those who have first committed to believe without the support of evidence. William James (1979) “The Will to Believe” in *The Will to Believe and Other Essays in Popular Philosophy*, Cambridge, MA: Harvard University Press.
adopt this sort of unquestioning allegiance to religious commitments. We should, however, recognize that some individuals do interpret their religious faith as having this kind of requirement, and I confess that it is easiest for me to understand this sort of commitment when it has religious justification. I discuss the issue, then, in the religious context, but do not intend it to apply only to those unquestioned commitments with religious foundations.

This objection is one that we must take seriously. It would be naïve to think that religious belief does not influence many deeply-held preferences and value commitments, and if that belief comes with the idea that questioning is itself a relinquishment of faith, then deliberative attempts to determine the adaptiveness of preferences may seem nearly as incapable as perfectionist accounts of showing secondary recognition respect to persons with such commitments. These people of faith would certainly see themselves as having competently arrived at a conception of the good, but deliberative accounts seem unable to recognize this competence, since such persons would be unwilling to do precisely what the deliberative accounts required for acceptable moral reasoning: recognize the weight of reasons that counted against their conception of the good, or offer reasons in its favor.

Unfortunately, this bullet may simply have to be bitten. There may be no good way of defending as non-adaptive some of those preferences that arise from intentionally unreflective faith. The concept of adaptive preferences is perhaps most often invoked in cases of women's endorsement of religiously-based norms that seem, from the outside, to be bad for them. If the motivation for the use of the concept is not to be entirely discarded, then, preferences of this kind cannot be entirely removed from scrutiny simply by virtue of a faith commitment that might itself have been unthinkingly instilled from a young age. I turn to these difficult issues in the next chapter.

Let us put the previous objection to the side, and assume, for the moment, that it is acceptable to expose persons' preferences to deliberation in order to determine whether or not they are adaptive. The question then arises of what role this deliberation is meant to play in determining what opportunities should be made available to the persons involved in the deliberation.

Sen has consistently emphasized the importance of freedom in human life, and has argued that the role of human development should be to expand the set of
freedoms that human beings can access. He has also, however, consistently argued that only the introduction of valuable opportunities expands a person’s freedom set, and that persons may come to have adaptive preferences that prevent them from accurately recognizing which freedoms are valuable. Deliberation, then, is meant to help to sort those things that persons have reasons to value from those that they do not, and this deliberation is meant, in turn, to inform the course of development. Even when discussion about values is abstract, the conclusions that come out of it will influence broader decisions about what values and opportunities are to be privileged in society.

Given the link between value formation and the values adopted by a group, then, one might worry that using a deliberative identification account of adaptive preferences will lead to a failure of primary recognition respect. This is effectively Robert Sugden’s worry about Sen’s proposal: Sugden argues that allowing questions about which opportunities persons have reason to value to be answered by a group will justify the paternalistic limitation of the opportunities that some group members would prefer to pursue. Sugden’s criticism of Sen is, I think, based on an uncharitable reading of Sen’s work, and I note Sen’s response below. It is worth considering Sugden’s criticism, however, because it raises a related issue that will be important to accounts of adaptive preference.

To understand Sugden’s criticism, consider an example. Imagine that the women in a community come together to deliberate about whether they have reason to value wearing the burqa. Current preferences to wear the burqa have developed under the influence of community norms that require it, so if it is found to be a practice that has no value, the preference to wear it will have counted as adaptive. In the best case, after offering each other reasons both in favor of and against the practice, the women will come to a unanimous decision. Whether they decide in favor of or against the practice, a unanimous decision would mean that critical reasoning had succeeded: reasons were exchanged, and minds were changed on the basis of those reasons. Adaptive preferences and otherwise problematic preferences (if any existed) were rejected, and nonadaptive preferences were either adopted or maintained.

---


See note 113 above.

The problem, Sugden notes, is that we are not entitled to assume that critical reasoning will always lead to a consensus about value. And where there is no consensus, he argues, it is unacceptably paternalistic for the majority to decide on the minority’s behalf what they have reason to value and which opportunities they ought therefore to have access to. Yet, he continues, this is precisely what Sen seems to be committed to arguing that the community should do. Sugden writes

>"[I]n a regime of the kind proposed by Sen ... I am assured the opportunities to lead those kinds of lives that a majority of my fellow citizens, after reflective deliberation and open debate, judge to be valuable. I am also assured the opportunity to participate in the debate on equal terms with everyone else. What I am not assured is the opportunity to live whatever kind of life I desire, within the constraints imposed by other people’s having similar opportunities."

While it is not clear if he would reject the idea that preferences can be adaptive in the first place, Sugden rejects the idea that adaptive preferences should be treated differently from other preferences. When he criticizes Sen for allowing the majority to reach a supposedly authoritative value conclusion that “is not shared by those people whose adaptive desires he tells us not to take at face value,” Sugden is disdainful of the idea that the phenomenon of adaptive preferences could give anyone a genuine ground for questioning the preferences of the minority in the first place.

Sugden’s critique of Sen is based on an uncharitable reading of the latter’s work. The brunt of Sugden’s objection concerns the appropriate response to adaptive preferences, and Sen has the resources to respond to such criticisms. In response to Sugden’s paper, Sen notes that his account of what opportunities persons have reasons to value is not meant to give a complete account of value; rather, this account should be complemented by procedural freedoms that ensure that no majority can run roughshod over a minority. But while this kind of response may be able to ensure that a deliberative theory of adaptive preferences does not violate the requirements of primary recognition respect by licensing coercive paternalistic actions, it does not clarify the way in which adaptive preferences should be identified.

Sugden asks us to consider a situation in which deliberation amongst community members about what there is reason to value ends in disagreement. On the deliberative
theory of adaptive preferences that I have discussed, such a situation raises a worry more germane to the problem of identification: which, if any, of the preferences in such a situation should count as nonadaptive? We could easily have a situation in which a group engages in deliberation about value and emerges divided. Some of the women involved could prefer to wear the burqa after some level of critical reflection, while the majority could have concluded that there was, in fact, no reason to value the practice. In such a situation, there arises a very real question about how an external observer should judge the adaptiveness of the preferences involved: should they defer to the reflective self-regarding preferences of the minority group of women, counting them as nonadaptive now that they have been subjected to some measure of reflection – or should they rely on the larger group’s reasoned decision that the opportunity in question is not one that should be valued? The problem is that there always remains the possibility that, upon further reflection, even the minority of women would overturn their own conclusion. And if most women in the community, or many women in similar situations, had already changed their minds, then an observer might think that the preferences of the women whose minds remained unchanged simply require more interrogation. The question of identification, then, remains: When should we actually count a preference as nonadaptive on a deliberative account? How much deliberation is enough? And when disagreement exists after deliberation, whose reasoned judgment should be trusted?

An empirical weakness: deliberation may not root out adaptiveness in practice

This brings us to an empirical weakness. As most deliberative democracy theorists recognize, deliberative practice will not always live up to deliberative ideals. When deliberating parties lack the tools, the willingness, or the knowledge to engage with one another and the issues, the benefits of deliberation will be compromised. Instead of improving reasoning and leading to more just outcomes for all, deliberation that takes place in less than ideal conditions can fall prey to rhetorical manipulation, insufficient

---

143 I present the example in this way for simplicity’s sake. Some women could also obviously think that the practice has value without wanting to engage in it themselves.

criticism of current practices, polarization of deliberators, or the domination of the agenda by more assertive or powerful individuals.

The problems that come with undertaking deliberation in practice are recognized by most proponents of deliberation, and there are a variety of proposals for minimizing those problems. Ackerly suggests “deliberation in enclaves,” which is meant to provide less powerful members of communities with a space where they can initially deliberate together without the potentially overwhelming influence of the community’s more powerful members.145 For his part, Archon Fung proposes the use of non-deliberative stop-gap measures to bring about conditions more conducive to deliberation.146 Carol Pateman proposes a state-sponsored basic income as a way of furthering democratization by ensuring that “individual freedom and citizenship are of equal value for everyone.”147 Noting that public dialogue in places like India can be “hard to achieve... despite democracy, because of low levels of elementary education, especially for women”,148 Sen proposes a requirement of compulsory basic education for deliberators that aims to ensure that all persons involved have at least some level of information and skill.149 And Severine Deneulin recommends incorporating a substantive conception of human good into the deliberation process in order to double-check that the exercise of deliberation is benefitting deliberators and allowing them to live more complete human lives.150 All of these theorists agree that deliberation requires some additional mechanisms or safeguards to ensure that it works as it is meant to.151

One such proposal has special relevance for the identification of adaptive preference: the introduction of a distant critical voice into a community’s deliberation.

151 Strangely, Mozaffer Qizilbash recognizes the practical problems that instances of adaptive preferences may pose for Sen’s work on deliberation, but prescribes more theory as a solution. He outlines an account of ideal deliberation in which preferences are autonomously formed through the power of reason, and power hierarchies are neutralized – and then notes that “[i]f this view of ideal deliberation supplemented Sen’s thick view, it would go some way to dealing with the standard problems associated with adaptive preferences.” Mozaffer Qizilbash (2007) “Social Choices and Individual Capabilities” Politics, Philosophy & Economics, 6(2), p. 181. Such ideal conditions can surely solve the problem of adaptive preference in theory – but in ideal conditions, adaptive preferences would not exist in the first place. Precisely what is at issue is the problem of how to attend to adaptive preferences in practice when ideal conditions are absent.
Ackerly and Sen both make such a proposal. Each of them is concerned with how communities ought to make collective decisions, and each recommends deliberation as a way of identifying the opportunities that persons have reason to value as a first step towards making those opportunities available. But there are different reasons to value having an opportunity: the fact that a person does value an opportunity is one kind of reason, but the fact that she might value it in the future, or that she rationally ought to value it provide other strong reasons to think that an opportunity is a broadly valuable one.\textsuperscript{152} And these latter kinds of reasons, unfortunately, can sometimes be difficult to access from one’s own current perspective. Indeed, both authors note, when a preference or value is broadly accepted in a community, those people who share it may be badly placed to recognize its limitations, and deliberation about it might lead to renewed endorsement without doing much to ensure that participants critically engage with the values at stake.

Recognizing this, Ackerly and Sen offer similar responses: they propose that critical outsiders should be given some role to play in local discussion about value. The hope is that such outsiders can serve as advocates for the kinds of opportunities that community members with more-or-less common perspectives may not champion on their own. Ackerly argues that the content of a person’s criticism is more important than the identity of the person who gives it,\textsuperscript{153} but notes that what she calls “multi-sited critics” have a special role to play in deliberation about value, since their positioning between worlds has forced them to become self-critical and self-reflective in a way that can help them to introduce a critical and reflective voice into debate with others who may previously have turned a less critical eye to their own circumstances.\textsuperscript{154} This proposal mirrors one made earlier by David A. Clark, who recognizes that what he calls cultural “insiders” may be oblivious to relevant ethical elements of their own practices “precisely because they are so omnipresent.”\textsuperscript{155} He accordingly argues that “outsiders” should be included in local debate about value since they may be better placed to “see and reveal things that an insider misses” or “inject new and sometimes needed ideas into an alien group’s development


\textsuperscript{154} Ibid., Chapter 5, esp. p. 156.

deliberations." And Sen draws on Adam Smith to illustrate the way in which entertaining views from "a certain distance" plays an indispensable role in helping moral reasoners "to avoid parochial prejudices and to examine a broader range of counter-arguments."\textsuperscript{157}

Such critical distance is especially important for adaptive preferences. Since adaptive preferences are adapted to limiting circumstances, a critical voice can help to introduce and advocate for the kinds of alternatives that may not be readily available in a local context. However, in the remainder of this section, I offer reasons to think that this mechanism alone may not be adequate for adaptive preference identification.

Before I continue, it is important to note that the primary task of these theorists is not adaptive preference identification. Rather, it is to improve collective deliberation about courses of action. As such, the ability to actually facilitate action is an important part of each of their projects: both theorists share a commitment to the idea that when a difficult choice arises, "it is a choice that the people involved have to face and assess."\textsuperscript{158} And (perhaps especially in development contexts) these hard choices will be ones that must be considered in decidedly non-ideal circumstances. Mechanisms like the critical voice can improve the deliberative process even if they cannot perfect it, and in imperfect circumstances an imperfect method for collective decision-making may prove to be the best alternative available. But I will argue that the same imperfections in method that must be put aside for purposes of facilitating group decision-making in non-ideal circumstances, should not similarly be put aside when we focus on the task of adaptive preference identification.

Consider one of Sen's examples of the way in which a distant critical voice can help to improve deliberation about values and preferences. Sen notes that in places where norms exist against women pursuing advanced careers in the sciences, "it may require considerable independence of mind" to imagine the norm being different. But, he continues, "observations from other societies where women have more opportunities could confirm that women have the ability to do just as well as men in the pursuit of science, given the necessary opportunities and facilities."\textsuperscript{159} As a way of informing collective deliberation in real-world circumstances, such a voice might indeed be the most useful alternative available. Doubtless it would be better to provide the participants themselves with experience of such alternatives, or to provide each of

\textsuperscript{156} Ibid, p. 165.
\textsuperscript{158} Sen (1999) "Development as Freedom" p. 51, emphasis added.
them with a source of income that would allow them to make more independent decisions, or a broad education that would make their decisions more informed – but in real conditions of inequality where these conversations must all too often be undertaken, such time- and resource-intensive measures will often be unavailable. The critical voice, then, goes some way towards providing these insights, even if the conditions for deliberation remain less than ideal.

If preferences were entirely rational, and adaptive preferences were merely preferences developed in the absence of knowledge of alternatives, then deliberation that included critical distant voices might also be enough to transform adaptive preferences into nonadaptive ones. But our preferences are not entirely rational, and we should not understand adaptive preferences as mere absences of the information that better alternatives exist. If this were the problem, then cases of adaptive preferences would be vanishingly rare. In our globalized world, very few people lack knowledge of other cultures so completely that they would actually\emph{fail} to realize that women in other countries work as scientists or hold political power. Indeed, very few people will even need to look beyond the limits of their own societies to find at least some role models. Behazir Bhutto, Aung San Suu Kyi, and Indira Gandhi offer prominent examples of politically powerful women in some of the places in the world where gender norms are most constraining; in the context of Sen’s example of the sciences, it is worth noting that the largest bio-technology company in India was founded and continues to be run by a woman: Kiran Mazumdar-Shaw.\footnote{The New Yorker (January 2, 2012) “Drug Test”.} So adaptive preferences cannot be preferences formed with a simple lack of knowledge that can be chased off by the introduction of a distant critical voice. In the remainder of the chapter, I consider some of the other mechanisms by which adaptive preferences might take hold, and which may not be so easily loosened by the introduction of a critical voice.

\textbf{Implicit bias}

When people engage in deliberation, they are mostly offering each other reasons. One party offers a proposition, and other parties reflect on the proposition and offer reasons either in favor of or against it. These sorts of reasons represent persons’\textit{explicit} attitudes: the ones that they reflectively endorse in light of evidence. But people are not fully rational creatures, and not all of our attitudes are either reflective or based on
reasoned evidence. And indeed, the attitudes that we hold explicitly can often come apart from those that we hold implicitly. We have probably all experienced this sort of situation before: I might instinctively dislike a person without being able to point to any negative fact about her, or I might continue to feel uncertain about a course of action that seems to have all of the evidence in its favor. Although I can clearly reason my way to a conclusion in both cases, some fundamental part of me continues to doubt that it is the right one.

We should be worried, however, not only about the fact that our implicit and explicit attitudes can come apart, but about the fact that our implicit attitudes may express significant bias. Indeed, our unreasoned implicit attitudes are likely to reflect long-standing prejudices against members of marginalized groups even when our explicit attitudes do not. Remember that I argued in the Introduction that we ought to be concerned with adaptive preferences of all sorts, from those that explicitly endorsed a value in all circumstances, to those that represented compromises with reality and intentions to act in limited circumstances that persons might rather not make or form in a better world. Accordingly, when we think about adaptive preferences, we should be concerned not only with the explicit attitudes that relate to the values that we endorse unconditionally, or the options that we explicitly rank against others. In what follows, I will show that we should also be concerned with the implicit attitudes and biases that resist the reasoned analysis that our explicit attitudes might respond to, while still affecting the way that we interact with the world and the chances we are willing to take.

Implicit bias affects what we do

We should worry about implicit bias first because it can sometimes be a better indicator of behavior than explicit attitudes. For instance, when implicit biases operate, subjects are more likely to identify the ambiguous actions of a black person as negative rather than neutral. Indeed, being primed to be aware of a person’s race can even lead to biased judgments in unrelated situations – when first shown a black face, subjects have become significantly more inclined to identify a harmless object as a

---

161 For a discussion of race and implicit bias, see Daniel Kelly and Erica Roedder (2008) “Racial Cognition and the Ethics of Implicit Bias” *Philosophy Compass*, 3(3). It is worth noting that while these sorts of biases are widely shared, implicit and explicit attitudes are more likely to come apart when persons explicitly reject the prejudice in question. Brian A. Nosek (2007) “Implicit-Explicit Relations” *Current Directions in Psychological Science*, 16(2), p. 67.

gun. And worrisomely, these tendencies can affect what we do even in cases where we do not endorse our biased behavior. The tendency to identify an object as a gun, for instance, is correlated more closely with measures of implicit bias than explicit attitude, and occurs even in persons who explicitly disavow racial bias. And in a separate case, employers who listed themselves as "equal opportunity" were found to be just as likely as those who did not to judge the same fictional CV more favorably when it was attached to a "white" name than when it was attached to a "black" one. If implicit biases are at times better indicators than explicit attitudes of the ways in which persons actually live, then we should be as worried about this kind of adaptive preference as we are about adaptive preferences expressed as explicit attitudes.

**Implicit bias can be self-directed**

Second, we should worry about implicit bias because it can be directed against members of one's own group. In the case of weapon bias noted above, black subjects were no less likely than white subjects to judge the harmless object to be a gun when it was shown to them after the image of the black face. Similarly, in another case of CV evaluation, female academics were just as likely as male academics to judge an identical CV more harshly when it was attached to a woman's name than when it was attached to a man's. Additionally, both men and women in the profession were more likely to vote to hire a supposed male candidate than a female candidate. And when tested, both women and men held negative implicit attitudes towards women in positions of authority.

This fact, taken together with the fact that implicit biases can influence behavior, should be particularly worrisome in the context of a discussion of adaptive preferences. Taking these two facts about implicit bias together, there is reason to worry that members of marginalized groups may not only fail to protest their constrained circumstances, but – as most theorists of adaptive preferences fear – may...
also actively, if unintentionally, participate in maintaining the circumstances that lead
to such potential adaptive preferences. If a woman implicitly holds that women ought
to behave deferentially towards men, then she may unconsciously reward other women
for deferring and sanction them for standing up for themselves – and she may do this
even if deliberation about the issue has led her to explicitly reject the problematic
norm, and count it as adaptive!

**Implicit biases and stereotype threat**

Stereotype threat offers another potential reason to worry about implicit bias in
deliberation. Stereotype threat is the result of the same prejudices that are reflected in
cases of implicit bias:

>[N]egative stereotypes are in part responsible for the under-performance of

minority members in stereotype-relevant domains. More specifically, those tasks
for which negative association exists between the task domain and the minority
group will represent a threat for minority group members; their preoccupation
with inadvertently confirming the stereotype will in turn lead to a decrease in
performance.168

For instance, women or other minorities who are conscious of their minority status
may perform poorly on math tests or tests of quantitative skill where they are
expected to do poorly.

Those who advocate deliberative accounts of adaptive preference identification
should be worried about the possibility of stereotype threat. Stereotype threat may
arise as a result of the anxiety that accompanies the prospect that one will re-confirm
the stereotype in question, but persons who hold implicit biases about members of
their own group that accord with stereotypes may be at even greater risk of under-
performing. Stereotypes that cast women as comparatively poor reasoners or bad
decision-makers are common in both the developed and developing world, and it is
plausible that women who engage in deliberation about shared communal values may
experience stereotype threat in a way that compromises their ability to present their
arguments clearly to other deliberators, or even to themselves.

**Implicit bias and experience**

The ways in which implicit biases change offers another reason to think that
deliberative accounts of adaptive preference identification may not be sufficient. In

“Changing the Ideology and Culture of Philosophy: Not by Reason (Alone)” Hypatia, 23(2).
short, implicit biases are tenacious. At the extreme, some studies suggest that implicit biases and attitudes may be impervious to all types of attempts at change.\textsuperscript{169} Most studies, however, agree that change, while possible, is the result of experience rather than reason. More specifically -- and more interestingly for the purposes at hand -- when individuals hold implicit biases about members of their own group, one of the most effective bias-reducing types of experience seems to be exposure to role models who buck the trend that the bias picks out. Some studies have shown that women in social contexts that exposed them to female leaders were less likely to hold stereotypic implicit attitudes about women as a group, and that they became increasingly unlikely to do so as the frequency of their exposure to female leaders increased.\textsuperscript{170} Other studies have had similar findings, but have noted that exposure to role models need not occur in person as long as the role models were highly visible and their successes were current.\textsuperscript{171} And even when broad implicit attitudes about one's own group do not change, exposure to role models can increase the self-efficacy of individuals in the relevant domain and encourage them to pursue further engagement in that domain.\textsuperscript{172} This means that individuals are likely to benefit from exposure to strong role models even if it is indeed the case that implicit attitudes resist all efforts at change. The common strand among these studies however, relates to individuals' identification with their role models: when individuals identify more closely with their role models, exposure to those role models will do more to reduce implicit bias and increase self-efficacy.\textsuperscript{173}

\textsuperscript{172} Jane G. Stout, Nilanjana Dasgupta, Matthew Hunsinger, and Melissa A. McManus (2011) "STEMing the Tide: Using Ingroup Experts to Inoculate Women's Self-Concept in Science, Technology, Engineering, and Mathematics (STEM)" \textit{Journal of Personality and Social Psychology}, 100(2). The results, however, are mixed. Another study found that women with female role models were more likely to continue to some male-dominated academic fields, but not in others. See Eric P. Bettinger and Bridget Terry Long (2005) "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students" \textit{The American Economic Review}, 95(2).
\textsuperscript{173} See also Penelope Lockwood (2006) "Someone Like Me Can Be Successful: Do College Students Need Same Gender Role Models?" \textit{Psychology of Women Quarterly}, 30.
Problems for adaptive preference

All of this should explain why simple knowledge of alternative possibilities, of the type in the female scientist example, may not be enough to encourage those with adaptive preferences to reject those preferences or identify them as problematic. Although Indian women are likely to be aware of women like Indira Gandhi and Kiran Mazumdar-Shaw, those whose backgrounds and experiences have been very different will be unlikely to identify with them. In these cases, efforts to deliberatively engage the potentially adaptive preferences of women living in fairly homogeneous circumstances will require role models who are closer to home. And if exposure to such role models is going to have the best chance of challenging existing preferences, it should be prolonged: ideally, they should be individuals with whom those with potentially adaptive preferences interact every day.

This discussion of implicit attitudes and biases suggests that the mechanism of included distant voices will not be sufficient for the identification of adaptive preferences. While it is certainly important that alternative possibilities be represented in the discussions that determine the adaptiveness of preferences, these possibilities should take the form of actual exposure to role models living in alternative ways, and these role models should be persons with whom members of the more local community can closely relate. But while advocates providing distant critical voices may be able to act as role models, they may not be involved in the community for any long period of time, and the distance of their social circumstances will often make it difficult for them to serve as persons to whom community members can relate.

There are some caveats to note here. The first is that most study of implicit bias and stereotype-threat has been undertaken in controlled experimental environments. It may be much more difficult to determine when and to what extent these concepts operate in real-world situations like the ones in which the adaptiveness of preferences would be judged. They may also operate differently in developed nations, where much of the research has been undertaken, and in developing nations, where adaptive

---

174 In this case, Ackerley's multi-sited critic may be especially valuable, since she has come from the community where she also serves as an outsider, and may therefore be easier to identify with.

175 These two factors – identification with and prolonged exposure to a role model – may perhaps be traded off against one another. Several studies have found that fertility rates go down and the status of women goes up in homes where more television is watched. The studies attribute the changes to the viewing of television programs that depict more egalitarian gender roles and smaller families than those found in local settings. Robert Jensen and Emily Oster (2006) "The Power of TV: Cable Television and Women's Status in India" The Quarterly Journal of Economics, 124(3); Eliana La Ferrara, Alberto Chong and Suzanne Duryea (2008) "Soap Operas and Fertility: Evidence from Brazil" Working Paper 633, Inter-American Development Bank.
preferences will also need to be identified. Nevertheless, several of the studies cited have attempted to investigate the phenomena outside of the laboratory setting, and they suggest that the value of long term exposure to role models holds in the real world as well as it does in laboratory settings. In any case, these studies suggest that further attention should be paid to the ways in which implicit biases and stereotype threat operate, and the behavior that springs from them. Similarly, the psychological literature on stereotype threat, at least, includes proposals for combatting the phenomenon that could perhaps be promisingly revised to reduce the threat in deliberative instances of adaptive preference identification, and this area also deserves more research.

A second caveat relates to the relationship between implicit bias and stereotype threat on the one hand, and the identification of adaptive preference on the other. In short, we may think that the information just recounted should not seem to be particularly disturbing in the context of adaptive preference identification. After all, if it is deliberation that is meant to determine the adaptiveness of a preference, then those trying to identify adaptive preferences should be more concerned with the explicit attitudes that are open to change through reason than with the implicit attitudes that seem likely to remain unchanged after deliberation. If a course of reasoned dialogue leads to the rejection of some preference that is now recognized to be adaptive, then the implicit biases that continue to support that attitude might seem more like an unfortunate residual consequence of the adaptive preference having been held rather than an obstacle to that preference having been identified in the first place. On this line of reasoning, implicit biases and the behavior to which they give rise will certainly count as problematic in their own right, but they should not count as obstacles to the courses of action that will lead to the kinds of role models required. After all, if individuals in a community have rationally agreed that a particular previously-held preference is adaptive and ought to be rejected, then they should agree to the courses of action required to change the corresponding non-rational implicit biases and attitudes.


But while this may be so in the best cases, we should not always rely on deliberation to weed out the preferences that ought to be counted as adaptive. At this point, I want to offer three reasons to think that some preferences that survive deliberation should still meaningfully be called adaptive, even if they do not amount to explicit and complete endorsement of the value concerned.

**Self-efficacy and the adoption of goals**

The first reason follows directly on from the previous discussion. When we talk about adaptive preferences, we are generally talking about the preferences that govern the lives that people actually lead. So when we consider proposals for the identification of adaptive preferences, we should understand the subject matter accordingly: the preferences that are being tested for adaptiveness will be actual preferences about lives lived by real people. On a deliberative account, then, it is these kinds of preferences rather than (or at the least, in addition to) broad questions of values that will be up for deliberation and consequent rejection or affirmation. And although implicit biases may not directly influence explicit attitudes in these areas, since the latter respond to reason in a way that implicit attitudes do not, I argue that we should understand implicit biases to have an important indirect effect on individual’s explicit attitudes about the ways in which they ought to live their lives.

When individuals adopt explicit goals for themselves, perhaps the most important thing to consider is their own perceived self-efficacy: that is, what they take themselves to be plausibly able to achieve. Indeed, according to some psychologists,

> Whatever other factors serve as guides and motivators, they are rooted in the core belief that one has the power to produce desired effects; otherwise one has little incentive to act or persevere in the face of difficulties.\(^{178}\)

In other words, a highly subjective matter – that is, self-efficacy – has a foundational role to play in the determination of an individual’s otherwise reason-responsive set of goals. To be sure, self-efficacy, while highly subjective, is not entirely unresponsive to reason. Instead, it can be influenced by explicit external feedback in particular areas.\(^{179}\)

But much of self-efficacy is the function of the sorts of implicit biases and attitudes that we have been discussing. For instance, even when male and female children perform equally well on academic achievement tests, female children have higher perceived


self-efficacy for social roles that are traditionally female than those that are traditionally male \(^{180}\) — yet may not experience lowered self-efficacy when “traditionally male” skills like mathematical reasoning are embedded in stereotypically feminine tasks.\(^{181}\) In these cases, children’s perceived self-efficacy responds not to the available relevant evidence, but to the broader gendered norms that surround them.

As long as adaptive preferences are deliberated about in the concrete rather than the abstract, then, we should treat as suspect any method that holds that deliberation will be sufficient to determine the adaptiveness of a preference. Consider the women in Sen’s example who are initially unlikely to take up careers in the sciences. If they have explicit beliefs that women can never be scientists, then those explicit beliefs may be affected by the presentation of the information that he proposes. Their own self-efficacy, however, is unlikely to be affected by the knowledge that women quite unlike them have succeeded in a scientific career in circumstances with which they are not familiar. Each individual woman, then, might accept the reasoning that women can be scientists, but reject the idea that she herself, or those like her, are likely to succeed in that area. And on these grounds, the women could collectively and reasonably endorse their current preference to steer away from science.

I think, however, that we should reject the idea that such a preference should be called nonadaptive. It is entirely relevant to the political project: it is the function of limiting circumstances and it disadvantages the parties who hold it. Yet in order to call it adaptive, we need some justification beyond its ability to survive reasoned scrutiny.

**Adaptive preference and identity**

The identities that persons take on when they deliberate should give us another reason to think that some preferences that survive deliberation should still meaningfully be called adaptive. It is uncontroversial to say that facts about particular persons affect the kinds of reasons that they are likely to take most seriously. Some of these facts will be related to a person’s particular interests: if I have a special interest in the environment, then the effects of overgrazing will be an especially compelling reason for me to oppose cattle ranching. Other facts will be related to a person’s social role: if

---


I am the parent of young children, I may be more concerned than others around me with the provision of quality childcare.

This is important for our discussion of adaptive preferences because the identities of the particular people involved in deliberation are likely to have a significant effect on the preferences and courses of action that are endorsed as reasonable post-deliberation. This fact is not problematic in and of itself. Communities, like persons, are different, and we should expect them to come to have different priorities and to come to different conclusions about the good. If we recognize anything like the fact of reasonable pluralism, then we should see the diversity of ways of living that persons might reasonably adopt as the appropriate result of free deliberation, rather than as evidence of its failure.

The problem relates to the fact that individual persons have multiple identities. No one is only an environmentalist or only a mother; indeed, the same person may be both environmentalist and mother — and devout Catholic and lawyer and amateur historian and lover of gardening and many other things as well. But while this multiplicity of identities can make our lives rich and rewarding, it also means that some of our identities may come into conflict with one another. A norm or practice that I may benefit from or have reason to endorse in the context of one facet of my identity may also be one that I am harmed by or have reason to resist in the context of another facet.

When a potential adaptive preference up for deliberation is one about which many people in a group are internally conflicted, that preference may once again survive reasoned deliberation. Consider the case of women living in purdah in a Sufi Pirzada community in India. When these women leave the house, they are expected to wear heavy veils. On the one hand, the women think that the veils are hot, cumbersome, ugly, and limit their mobility and independence. But on the other hand, the women recognize the veil as a marker of their superiority to non-Pirzada Muslim women, and they see it as a symbol of modesty and propriety — traits that they value. Here, the same practice interacts with multiple facets of the women's identities, and

---


these facets come into conflict with each other. As female members of the Pirzada community, they have good reasons to wear the veil; as human beings engaged in the daily chores of living, they have significantly less reason to wear it.

On balance, we can imagine that these women might, after deliberation, continue to see themselves as having reason to value the practice of veiling. And such a decision, once reached, would have many reasons in its favor, and could be made rationally. The preference that it endorses, however, should not therefore be seen as nonadaptive. Once again, it seems to be a preference formed without real alternatives and, as the women themselves recognize, it disadvantages them in at least some important ways. While the rationality of their decision requires us to show the women both primary and secondary recognition respect, it should not cause us to ignore these other elements of adaptiveness. Yet if we can imagine such a preference being maintained after deliberation, recognizing these elements of adaptiveness will require some method beyond deliberation for the identification of adaptive preference.

**Deliberation and inequality**

Finally, the conditions of deliberation themselves might lead to adaptive preferences being endorsed at the communal level even when they are rejected by some at the individual level. As we saw, the theoretical results of deliberation may be much more attractive than the results that deliberation produces in practice. This is because the conditions necessary to facilitate the ideal are fairly demanding. In an important essay on deliberative democracy, Joshua Cohen notes that the deliberative democratic ideal “involves public deliberation focused on the common good, requires some form of manifest equality among citizens, and shapes the identity and interests of citizens in ways that contribute to the formation of a public conception of common good.” In addition, it requires a “willingness to revise preferences in light of discussion, new information, and claims made by fellow citizens.” But of course, in practice, any of these conditions can be absent: participants may be more concerned with their own good than the common good, they may be unwilling to revise their preferences in light of new evidence or the claims of others, or they may “deliberate” in conditions of inequality that either intentionally or unintentionally lead to the effective silencing of some voices.

---


Given that the identification of adaptive preferences is supposed to come about through a process of real deliberation, we should recognize that the process may be compromised by shortfalls in any of these areas. Individuals with a vested interest in a particular practice may see it as disadvantaging other members of the community, but may nevertheless try to convince those disadvantaged members that it benefits them. And if those individuals are disproportionately powerful, they may use their authority as an additional means of convincing their fellows that the preference in question should not be counted as adaptive. Indeed, even well-intentioned powerful individuals may skew the discussion if their perceived authority means that their comments are given more weight than the comments of the less powerful individuals who bear the brunt of the practices in question. And these effects may be exacerbated if the implicit biases or low self-efficacy of less powerful deliberators support the positions that more powerful deliberators defend. In this way as well, then, deliberation can yield group endorsement of a preference that ultimately deserves to be counted as adaptive.

Conclusion
The tool of deliberation has much to offer to a discussion of adaptive preferences, and many of the issues that lead to caution in its use may have either whole or partial remedies. Each of the discussed areas of potential shortfall from the deliberative ideal can and should be attended to individually. Insofar as solutions in each of these areas are found, deliberation will serve as a better tool for both the broad task of communal decision-making and the narrower task of adaptive preference identification. As long as the hazards noted in this chapter persist, however, an account of adaptive preferences should include a mechanism that ensures that the narrower task of adaptive preference identification can be met. Accordingly, I offer an additional condition for a desirable account of adaptive preferences:

Consideration 9: Mechanism for "Double-checking." An account of adaptive preferences should utilize some mechanism for ensuring that it does not yield dangerously parochial or conservative results.
Chapter Three: Children and Adaptive Preferences

Consideration of children has been almost entirely absent from discussions of adaptive preferences. This is peculiar, not only because children are ubiquitous to every social situation, but also because the ways in which we think about and take ourselves to permissibly treat children are often so different from the ways that we allow ourselves to think about and treat adults. In this chapter, I will investigate some of the consequences that consideration of children will have for my project, and conclude that appropriate responses to adaptive preferences are different when those preferences currently occur in adults than when they might occur in children.

Adaptive preferences and primary recognition respect

In Chapter One I developed a framework for thinking about the kinds of respect that persons are owed. There I focused primarily on secondary recognition respect, considering whether perfectionist accounts of adaptive preferences could show this kind of respect for persons. At this point, I want to return to the issue of primary recognition respect, in order to consider more carefully what it might mean for an account of adaptive preferences to show persons respect of this kind. Remember that showing persons primary recognition respect meant showing appropriate consideration for them as moral agents for whom the ability to live a life in accordance with their own convictions has great value. In other words, it concerned the extent to which a person can succeed in directing the course of her life.

But what precisely does this mean? Showing such consideration could mean two very different things. Here, Isaiah Berlin's well-known discussion of positive and negative liberty will be informative. In Berlin's account, negative freedom is freedom from coercion and interference. In his words, "if I am prevented by other persons from doing what I want I am to that degree unfree." Here, the emphasis is on persons' ability to act on the convictions that they actually have: if we think that the ability to live a life in accordance with one's own convictions has great value, the argument for negative freedom goes, then persons must be permitted to live in

---

accordance with the convictions that they have. And as critics of the concept of adaptive preferences point out, adaptive preferences will at least sometimes count as convictions: as Uma Narayan notes of such preferences, “despite undeniable distortions, these are in fact the values, attitudes, and choices that define for these women the lives they currently have and value, and the selves they currently are and in many ways want to remain.” So understanding primary recognition respect in terms of negative freedom recommends treating adaptive preferences like any other preferences. It means protecting adaptive preferences from coercive paternalistic intervention to the same extent that any other preferences would be protected.

Positive freedom, on the other hand, is not a freedom from, but a freedom to: it is a freedom to determine what it is that I want, not merely as a reaction to my circumstances, but in light of my own conscious purposes. This is the conception of freedom that most proponents of the concept of adaptive preferences are likely to endorse. They hold that unjust or otherwise problematic circumstances can cause persons to alter or constrain their values and preferences in ways that they would not do in better circumstances, and that such values and preferences are accordingly not genuinely a person’s own. For them, then, recognizing the interest that persons have in living their lives in accordance with their own convictions means helping persons to develop preferences that are not so constrained. Here, it will seem disrespectful to treat as decisive preferences that support the replication of the problematic limiting conditions under which adaptive preferences were themselves formed. So the emphasis here is on the preferences that a person might have, rather than preferences that she actually does; on this account, we can understand primary recognition respect as a facilitation of the preferences that a person would develop on her own if she were in less limiting circumstances. When primary recognition respect is understood in terms of positive freedom, showing that respect points towards expanding the options of persons with adaptive preferences even when they initially prefer their problematically limited circumstances.

---


There are certainly complications involved in understanding primary recognition respect in terms of positive freedom. How do we know when persons have preferences that they would not have otherwise developed? Does showing primary recognition respect mean pushing persons to develop certain preferences that we suspect that they would have developed otherwise? How do we go about determining which preferences they would have developed otherwise?

Berlin himself is rightly concerned with these questions, and it is on their basis that he holds that it should be the job of the state to protect negative rather than positive freedom. If we are willing to compromise another's negative freedom for the good of her positive freedom, Berlin warns, we can only do so on the grounds that we know better than she what is in her interests – that is, what she would choose, or prefer, if she were positively free. And this, he says, has unacceptable consequences:

Once I take this view, I am in a position to ignore the actual wishes of men or societies, to bully, oppress, torture them in the name, and on behalf, of their 'real' selves.\textsuperscript{190}

So given the possibility of paternalistic coercion when the demands of negative and positive freedom conflict, Berlin rejects the idea that the state should encourage positive freedom in its citizens.

This worry is a cogent one, and the negative freedom theorist has a good \textit{prima facie} case for rejecting state support of positive freedom on its basis. What the negative freedom theorist is not entitled to do on its basis, however, is reject the concept of positive freedom entirely. Persons can be wrong about what they want and what is good for them, and can discover that their previous convictions were wrong or mistaken by their own lights. If I am an emotional eater, food is not really what I need; instead, I need (for instance) a friend who can help me to feel less lonely and depressed. Similarly, if I leave an abusive relationship, and come to realize that it was abusive only later when I enter into a better relationship, I will likely come to see that a non-abusive relationship would have been better for me all along. But when I recognize these things, I am not simply changing my mind about what is good for me – I am realizing that I was wrong to turn to food, or to see my initial relationship as valuable, in the first place.

Some negative freedom theorists may simply deny this, claiming that I merely change my mind in cases like these, and that both the original and the revised options

\textsuperscript{190} Berlin (1959) \textit{Two Concepts of Liberty}, p. 18.
were valuable as long as I valued them. There is little that I can say to these people. Instead, I aim the arguments in the rest of this chapter – and indeed, in the rest of this thesis – at those who find it plausible that I can be wrong about my interests in such cases, whether or not they think that being wrong in these ways warrants any kind of intervention into my life to protect my positive freedom.

**Autonomy and children**

At this point, I want to begin to think of these issues in terms of autonomy. Clearly, positive freedom and autonomy are not identical. To begin with, autonomy can be associated with negative freedom rather than positive freedom, as it is when it is attached to the right to act autonomously.\(^2\) When autonomy is thought of as a capacity, however, the two concepts have much in common. There is little agreement about the best way to understand the capacity of autonomy, but virtually all accounts rely on a kind of self-rule that requires a fairly demanding type of engagement with and endorsement of oneself and one’s values.\(^3\) Characterizing the thread that binds most accounts of autonomy together, John Christman and Joel Anderson write that the concept

finds its core meaning in the idea of being one’s own person, directed by considerations, desires, conditions, and characteristics that are not simply imposed externally on one, but are part of what can somehow be considered one’s authentic self.\(^4\)

This mimics Berlin’s idea of positive freedom as being “moved by reasons, by conscious purposes which are my own, not by causes which affect me, as it were, from the outside.”\(^5\) In the rest of the chapter, unless otherwise specified, I will use the term “autonomy” to refer to the sort of autonomy that overlaps with positive freedom, and I will use the two terms interchangeably.

The concept of adaptive preferences is intimately related to autonomy understood in this way, since adaptive preference theorists are concerned precisely

---

1. I explore the idea of autonomy as a right further in Chapter Four.
with those instances in which unacceptably limited circumstances do move people from the outside to come to endorse desires and conditions that they would not otherwise endorse. Adaptive preferences, however, are not simply the opposite of autonomy. Some accounts of adaptive preferences explicitly deny that adaptive preferences are autonomy deficits, and even I, who support an autonomy-based account of adaptive preferences, will argue that adaptive preferences are only a subset of all non-autonomous preferences.

The turn from adaptive preference to autonomy, at this point, is motivated by the need to expand discussion of the concept of adaptive preferences to an area which has previously been ignored: the unique needs of children. Most discussions of adaptive preferences, as we saw in the Introduction, focus on the ability of the concept to show respect for adults. But while this issue is certainly important, it is not the only issue at stake. Both the needs and the permissible ways of treating children have traditionally been understood to differ significantly from those of adults. Recognizing this is important, because although children are ubiquitous in every social situation, their needs are rarely taken as seriously as those of adults in philosophical discussions and are often not recognized at all. It is because there is no literature on children and adaptive preferences that I will focus in this chapter on the literature on children and autonomy in order to bring the needs of this too often invisible group into the broader discussion on adaptive preferences. Considering this literature will help us to understand not only what changes theories of adaptive preferences might need to make in order to accommodate children; it will also help to understand the relationship between adults and their own adaptive preferences more clearly.

As Joel Feinberg notes, there are two primary reasons that one might value autonomy: either because its exercise or presence is intrinsically important, or because it is instrumentally valuable in promoting persons' well-being or self-fulfillment. I will consider the first possibility in the next chapter, but for now I will focus on the second. In explanation of this second possibility, Feinberg writes that "a given normal adult is much more likely to know his [sic] own interests, talents, and natural dispositions... than is any other party, and much more capable therefore of directing


his own affairs to the end of his own good."\textsuperscript{197} This position is a fairly standard one — indeed, John Stuart Mill relied on it heavily in support of his famous harm principle.\textsuperscript{198}

But while this position is a common one, it is not always clear in what sense it is meant to count in favor of autonomy. As I mentioned above, autonomy can be associated either with positive or negative liberty. When it is associated with positive liberty, an attribution of autonomy makes a descriptive claim about the extent to which a person exhibits the traits of reflective self-mastery listed above. When it is associated with negative liberty, attribution of the term involves a normative claim about the right that a person has to make otherwise permissible decisions about her life without interference. It is often taken that the normative claim should be made on the basis of the descriptive one, but this need not always be the case, and in the next chapter I will argue that the possession of descriptive autonomy should not in fact be the trait upon which one's normative autonomy hinges.

For now, it is only important to note that it is not entirely clear whether Feinberg and Mill's points are meant to apply to all persons, or only to those who are descriptively autonomous. Leaving their conclusions aside, it is certainly the case that the possession of descriptive autonomy makes it more likely that a given person will know her own interests, talents, and dispositions better than a third party. To see why this is so, I will momentarily turn to the literature on children and autonomy, but I must first briefly discuss what I mean by a person's interests.

What constitutes a person's interests or well-being is a complicated question.\textsuperscript{199} I have already rejected the adoption of a perfectionist account of persons' interests, for reasons given in Chapter One. I also rejected extreme subjectivist accounts earlier in this chapter. Instead, I will assume an intermediate informed desire account of persons' interests. Peter Railton offers such an account. Here,

\textsuperscript{198} John Stuart Mill (1863) \textit{On Liberty}, Boston: Ticknor and Fields, p. 147.
An individual's good consists in what he would want himself to want, or to pursue, were he to contemplate his situation from a standpoint fully and vividly informed about himself and his circumstances, and entirely free of cognitive error or lapses of instrumental rationality.200

On this kind of account, we have both subjective and objective interests. Our subjective interests are those that we in fact take ourselves to have in our real lives where we are neither fully and vividly informed nor free of cognitive error, while our objective interests are those that we would take ourselves to have in the hypothetical and counter-factual alternative case.201 And clearly, on an informed desire account such as this one, our subjective and objective interests can come apart.

Having descriptive autonomy helps us to ensure that our objective and subjective interests line up as far as possible. This is widely recognized in the literature on children and autonomy. That literature focuses with remarkable frequency to the 1972 U.S. Supreme Court case Wisconsin vs. Yoder.202 In this case, the Court decided in favor of a group of Old Order Amish who challenged a Wisconsin state law that made education through age 16 compulsory, on the grounds that it violated their right to religious freedom and endangered the very existence of their community. The majority of theorists writing on children hold that this case was decided wrongly, often on the grounds that it violated the children's right of exit from their community. I want to focus, however, not on the broader question of whether children's exit rights are violated by the ruling, but on two factors that theorists writing on Yoder often point to as mechanisms by which that right is protected.203 I focus on these mechanisms because they accord with several influential accounts of autonomy offered in the literature,204 and because they offer a plausible account of what is required for the


positive freedom that allows children to form their lives on the basis of reasons which are meaningfully their own. They are also the mechanisms by which autonomy enables the alignment of objective and subjective interests.

The first mechanism required for autonomy is access to a set of alternative options. If children are unaware of other lifestyles and systems of values, it will be very difficult for them to consider these alternatives as possibilities that they might want to pursue. As Rob Reich notes, when a person has been “rigorously seclude from value diversity, his or her horizon of choice is severely limited... it is impossible to pursue an end or endorse a value that one does not know exists.” If children do not know what their options are, then they will be in no position to choose for themselves which options are most valuable.

But how many — and indeed, which — options does autonomy require? A child’s option set cannot be fully closed in even the most extreme circumstances. Even secluded Amish communities allow their youths to explore the outside world in a ritual called “rumspringa” before they choose to become adult members of that community. But a child’s option set cannot be fully open, either. Every child is born into a social circumstance, and every concrete circumstance will make some options more immediately available than others. In Claudia Mills’ words, it is simply impossible to raise children “without steering them, however imperceptibly, toward one option rather than another.” Given that many of a child’s early activities will be chosen by her parents, the path dependence of most lives means that activities chosen by someone else in childhood are likely to be the ones that persons continue to engage in and identify with.

If a child’s set of options can be neither fully closed nor fully open, then choosing autonomously will require possessing the tools to decide between the unavoidably limited set of options that she has, and to recognize the possibility that she might need to seek out others. Without these tools a child’s chance to direct her own life would be more formal than substantive: if I am given the formal right to leave the Amish community in which I have been raised, I will only have any real chance of taking advantage of this opportunity if I am able to weigh the value of staying against the

---


value of other alternatives with which I have little familiarity. When so many factors weigh in favor of staying in an insular community, the genuine possibility of leaving requires significant independence of mind. This, then, is the second important mechanism of autonomy: the possession of the independence of mind that allows her to interrogate the commitments that she has actually come to have and to consider the alternatives less obviously available to her. Autonomy conceived of in this way ensures that objective and subjective interests line up as far as possible, because it makes it more likely that people will become aware of the interests that they would endorse for themselves if they were fully informed, and it gives them the tools to consider carefully which options they should endorse.

Note, however, that promoting positive freedom for children in this way will sometimes involve impinging upon their negative freedom. If children in the Amish community were required to attend school until 16 as the state of Wisconsin initially required, then those children who wanted to drop out at 14 would be coercively prevented from doing so. Indeed, non-Amish children are coerced into school attendance every day in many countries, and most theorists and policy-makers think that requiring children to attain the skills that education provides is not only permissible, but obligatory.

Here, it is striking to note the difference in the treatment of children and adults. While mandatory school attendance for children could be justified on many grounds, including ensuring that children have the skills to support themselves and avoid becoming dependent on state resources as adults, the authors discussed above all argue for education on grounds related to the enhancement of autonomy, or positive freedom. And indeed, even when the other-regarding arguments about state resources hold true for adults (that is, even when adults lack very basic skills such as minimal

---

208 For discussion of some such factors, see Susan Moller Okin (2002) "Mistresses of Their Own Destiny: Group Rights, Gender, and Realistic Rights of Exit" *Ethics*, 112.

literacy that make employment much more likely), no theorists argue that those adults should be coerced into mandatory education.

To the contrary, when it comes to adults, many liberal theorists hold that the state is not even permitted, much less obliged, to use its resources to encourage adults to act autonomously. This position is grounded in John Rawls's broadly political liberal philosophy, which begins with what he calls the liberal principle of legitimacy. According to this principle, political power is "legitimate only when it is exercised in accordance with a constitution (written or unwritten) the essentials of which all citizens, as reasonable and rational, can endorse in the light of their common human reason." But given what he calls the fact of reasonable pluralism — that is, the fact that "a diversity of conflicting and irreconcilable yet reasonable comprehensive doctrines" is a permanent feature of democracy — the set of values that can satisfy the liberal principle of legitimacy will be of a very particular sort. Since citizens' comprehensive doctrines will conflict with one another, no single comprehensive doctrine can receive endorsement from all of them. Accordingly, when choosing a conception of justice to govern their shared state, they should practice a "method of avoidance" and adopt the only kind of doctrine that all citizens can endorse — that is, one that puts all metaphysical and otherwise comprehensive matters to the side and makes claims only in the realm of the political. And the importance of adopting nothing more than a political doctrine at the state level comes from worries about coercion: for Rawls, all political power is coercive power, so given the liberal principle of legitimacy and the fact of reasonable pluralism, the state cannot justifiably coerce citizens into acting in accordance with a conception of the good life that many of them cannot and do not share.

For most political liberals, even state commitment to the value of autonomy (at least beyond the political realm) is unacceptable on these grounds. Rawls does hold that the state may permissibly encourage citizens to see each other as possessing a limited form of "political autonomy", which refers to "the legal independence and assured integrity of citizens and their sharing equally with others in the exercise of political power." But this form of autonomy is extremely limited: it does not extend to

---

210 As we will see, a few of these theorists also deny that the state can use its resources to encourage a robust form of autonomy in children.
212 Rawls (2001) Justice as Fairness, p. 34.
encouraging in persons autonomy as a "certain way of life and reflection [characterized by] critically examining our deepest ends and ideals." To encourage this sort of ideal in citizens would be, in Charles Larmore's words, to have "betrayed in fact the liberal spirit." William Galston holds that such a betrayal would have consequences which might seriously curtail the negative liberty of some citizens. In his words,

The decision to throw state power behind the promotion of individual autonomy can weaken or undermine individuals and groups that do not and cannot organize their affairs in accordance with that principle without undermining the deepest sources of their identity.

Those political liberal theorists who write specifically on children, however, are less willing to accept that the state has no business encouraging positive freedom in its citizens. They generally want neither to compromise the negative freedom that political liberals claim for adults, nor to grant that negative freedom is all that children require. Of those in this camp, some try to argue for state encouragement of autonomy on grounds that would seem appealing to political liberals and broader liberals alike: specifically, they claim that possessing a degree of autonomy is a prerequisite for meaningfully exercising the rights and liberties that political liberals protect for citizens. Others actually hold that political liberalism itself necessarily sees citizens as having an interest in developing the "capacity of being able to reason in a critical and rational fashion about one's conception of the good." And indeed, even those who, in a seemingly political liberal manner, explicitly reject state endorsement of

---

215 John Rawls (1997/2005) "The Idea of Public Reason Revisited" in Political Liberalism: Expanded Edition, New York: Columbia University Press, p. 455-6. However, it is of course possible -- and perhaps even likely -- that many citizens would happen to come to develop this stronger form of autonomy through the acquisition of political autonomy. It is also worth noting that the society for which Rawls advocates would generally provide persons with the other requirement of autonomy: that is, exposure to a broad variety of alternatives.


218 For Amy Gutmann, the critical thinking skills that a child needs should depend upon the demands of the society in which she lives, and the commitment to critical reasoning involved therefore seems plausibly to be a political one. Gutmann (1980) "Children, Paternalism, and Education: A Liberal Argument". And although Richard Arneson and Ian Shapiro also mount an independent defense of children's interests in autonomy, their primary argument is that a degree of autonomy is required in order to exercise the rights that political liberals endorse. Arneson and Shapiro (1996) "Democratic Autonomy and Religious Freedom".

219 Andersson (2011) "Political Liberalism and the Interests of Children", p. 295. Ruth Abbey also mounts an internal critique of political liberalism that holds that individuals must have developed autonomy in their personal lives in order for political liberalism to respect their choices to live inegalitarian lives. See Ruth Abbey (2007) "Back Toward a Comprehensive Liberalism? Justice as Fairness, Gender, and Families" Political Theory, 35(1).
autonomy as an ideal seem to argue for something much thicker than the political autonomy that Rawls suggested.²²⁰

In this chapter, I will not try to establish that the internal requirements of political liberalism either require or allow the use of the state for the development of the internal capacities for autonomy. I only want to show here that the arguments of theorists who think specifically about children overwhelmingly support the position that being fair to the interests of children requires using the apparatus of the liberal state to ensure that persons develop a significant degree of autonomy. This need not require treating autonomy as an ideal of a good life²²¹ in the way that Larmore feared, but it does seem to require giving it a much greater place in a liberal society than political liberals would prefer.

Establishing a difference between children and adults

When children lack either a range of options from which to choose, or the independence of mind and critical thinking skills required to choose between them, their preferences run the risk of becoming adapted to the circumstances that they find themselves in. When those circumstances are socially problematic in whatever sense is relevant, then their preferences become adaptive in the normative sense that I have been using the concept in this thesis. The development of such preferences is precisely what adaptive preference theorists wish to avoid, and the broader literature on children rightly notes that there is no problem with using the coercive power of the state to ensure that children develop the kind of positive freedom that would allow them to develop preferences beyond their immediate, limited circumstances.

The political-liberal literature that concerns adults, however, does not hold that the coercive power of the state can be similarly extended to expand the positive freedom of adults. Critics of the concept of adaptive preferences similarly think that persons with adaptive preferences should not be forced to act against their own judgments, inclinations and preferences, even in cases where it might be better for

²²⁰ Directly after stating that “the state has no business promoting comprehensive philosophical ideals of individual autonomy” Stephen Macedo claims that “all children should have an education that provides them with the ability to make informed and independent decisions about how they want to lead their lives in our modern world.” This certainly seems to me to recommend a range of options and the criticalness of mind that allows persons to make their decisions independently of pressure from others. Stephen Macedo (2000) Diversity and Distrust: Civic Education in a Multicultural Democracy, Cambridge, MA: Harvard University Press, p. 207.

²²¹ For a discussion of autonomy as an instrumental value, see Arneson and Shapiro (1996) “Democratic Autonomy and Religious Freedom”.
them to do so. But why should this be? Is there a non-arbitrary difference between children and adults that makes this difference in treatment justified?

The assumption in the literature on children seems to be related to the possession of the descriptive type of autonomy that overlaps with positive freedom. Out of respect for the autonomy of adults, Feinberg holds, other parties must allow adults to act in ways that foreseeably harm their future interests because their "autonomy rights" robustly protect their liberty of choice.\textsuperscript{222} Children, however, "cannot very well exercise [their] free choice until later when [they are] more fully formed and capable,"\textsuperscript{223} and their rights must accordingly be protected for them until they are adults and can meaningfully – that is, autonomously – make decisions about their own lives.

Put into the terms of positive and negative liberty that I have been using, the justification looks like this: because adults already have positive freedom, their negative freedom must be respected. Conversely, since children do not yet have positive freedom, they are not yet owed negative freedom. Such a distinction would make sense. After all, it is the possession of positive freedom that enables persons to ensure that their objective and subjective interests line up. But why think that adults have positive freedom, or autonomy, while children do not? Whether a person possesses autonomy will necessarily depend upon what is meant by autonomy, and no account of autonomy will draw the line between children and adults in the right place. The definition of autonomy used above is in fact quite demanding, and while it is indeed unlikely to apply to most children, many adults may also fail to meet its criteria. Adults with serious mental disorders like paranoid schizophrenia and severe mental retardation will of course fail to count as autonomous, but so will many so-called "normal" adults who have been exposed to few options or have failed to develop independence of mind. Accordingly, their negative freedom will not be protected.

To make these adults count as autonomous, we could adopt a weaker account of autonomy such as Harry Frankfurt's: here, a person is autonomous just so far as her higher-order preferences endorse her lower-order ones.\textsuperscript{224} Here, we would count as descriptively autonomous the person who makes her later choices in accordance with her previous dispositions and inclinations. This account of autonomy is much less

\textsuperscript{222} Feinberg (1980) "The Child's Right to an Open Future", p. 78.
\textsuperscript{223} Ibid., p. 77. For similar support for the justification, see Timothy Michael Fowler (2011) "The Problems of Liberal Neutrality in Upbringing" Res Publica, 16.
\textsuperscript{224} Harry Frankfurt (1971) "Freedom of the Will and the Concept of a Person" The Journal of Philosophy, 68(1).
demanding than the previous one: it does not require a person to be able to critically examine the particulars of her situation – instead, it only requires her to be able to choose in accordance with them. And it does not require such a broad set of alternative options: those that are locally available are also likely to be those that are most in harmony with a person's previous dispositions. But the problem here is the converse of the previous one. While virtually all non-disabled adults will count as autonomous, many children will as well, and will accordingly be owed negative freedom.

Indeed, the fact that adults will not necessarily be more autonomous than children should be unsurprising given the current discussion of the way in which children's autonomy is promoted. If we recognize that children possess only a minimal degree of autonomy, we could adopt two possible accounts of how they become autonomous adults: either they do so automatically, or else their autonomy must be developed. If children automatically become autonomous adults, then it seems that the only responsibility of a liberal state should be to ensure that some range of options remains open to a child at the point at which she has become autonomous. But few of the theorists writing on children seem to think that providing a range of options is sufficient. Instead, they criticize the decision in *Wisconsin v. Yoder* in part because they worry that removing children from high school at 14 seriously jeopardizes their acquisition of the critical thinking skills that autonomy requires. So it seems much more plausible to say that individuals tend towards autonomy, but that this capacity for autonomy can fail to be developed.²²⁵

At this point, we are left with a choice: we can adopt a demanding account of autonomy and recognize that many adults will no longer be owed negative freedom, we can adopt a comparatively undemanding account of autonomy and recognize that many children will now be owed negative freedom, or we can adopt a new justification for owing persons negative freedom. I will take the third tack. Doing so is especially important given my concern with adaptive preferences. Insofar as there is something wrong with having adaptive preferences, children who do not yet have them should be protected, as far as possible, from developing them. This will involve pushing them to experience and consider new possibilities that there is reason to think are valuable, whether or not they want to. The same, however, should not be the case for adults.

²²⁵ Harry Adams makes this point clearly: "autonomy is not an attribute that all adults possess automatically or necessarily, simply in virtue of their being adults (or humans, or moral agents, or citizens, etc.); but rather, autonomy is an attribute that *some* adults come to possess as an ideal end product of a contingent process of personal development." Harry Adams (2008) *Justice for Children: Autonomy Development and the State*, Albany: State University of New York Press, p. 2.
Persons with adaptive preferences already find themselves in limited and disempowering circumstances, and pushing them against their will to experience a new set of circumstances to which they are averse would only further disempower them.

The new justification that I propose is this: that the objective interests that we have depend in part upon the histories that we have with what I will call our “commitment-objects.” Among the various interests that persons can have are those that I will call “commitment-interests.” I follow Cheshire Calhoun’s definition of a commitment as “both an intention to engage with something... and a preparedness to see to it that that intention to engage persists,” at least across a certain range of cases. I will distinguish, however, between two ways in which one may have an interest in commitments. First, one can have an interest in having some commitments rather than no commitments. Second, one can have an interest in the objects of particular commitments in a way that goes beyond the interest that she would have in that object if she were not committed to it: because her engagement with the object has given shape to her settled adult life, that object contributes more to her interests than other objects do. So, for example, if I have an interest in having some commitments in my life, then a commitment to environmental preservation may fill that role as well as any other. But having adopted this particular commitment, I will now have a stronger interest in the preservation of an endangered species or the passing of a cap-and-trade emissions bill than I will in many other worthy causes unrelated to the environment.

I take it to be uncontroversial that at least the great majority of adults have an interest in having some substantive commitments. It is hard to imagine an adult life having shape, meaning, or direction without an intentionally-maintained engagement with at least some values across the course of that life. Calhoun’s article aims to show that this will not be the case in every single circumstance, but she does not deny that in the vast majority of lives, commitments play an indispensable role. This is no less true for adults with adaptive preferences than it is for others. It is difficult, however, to determine just which commitments a particular person ought to have – and this is where the problem arises when it comes to adaptive preferences. Adaptive preference theorists think that, on at least some measure, persons with adaptive preferences have adopted the wrong set of preferences, values, or commitments. But in the remainder of the chapter, I will argue that the commitments which it is in a person’s objective

---

interests to have will depend only in part upon characteristics intrinsic to individual commitment-objects themselves: it will depend as well upon the value that comes from them being a part of a commitment that gives shape to a person's settled identity.

Consider the analogous case of friendship. It is uncontroversial both that we all have an interest in having friends, and that what makes a friend a good friend for me depends not only on the features of a friendship that generally count as good-making (loyalty, honesty, generosity, etc.), but also on features particular to my own life (a person interested in poetry may count as a good friend for me but not for you, if I, unlike you, am interested in poetry). In the latter cases, individuals presumably have privileged epistemic access to the sorts of friendships that would enrich their lives. The more interesting point, however, relates not to features of friendships that are good-making according to individual interests, but to those features that are generally understood to be bad-making (for instance, intolerance, impatience, rudeness). If I am choosing a friend from an array of strangers, these sorts of bad-making features are likely to count against a person – but having already befriended some person who has those features, they may actually come to contribute to the friendship's being good for me. I can come to sincerely love my friend for features that I previously saw as shortcomings, and I could genuinely regret it if such a character trait were lost at some later time. After all, this character trait is just part of who my friend is. Importantly, I might come to value a previously undesirable trait in a person no matter how I came to befriend her, whether I did so out of necessity (as I might if no other friend is available, or if we have known each other since we were in diapers) or because other good-making features outweigh the bad. And in such cases, outsiders will almost certainly be badly placed to determine whether a particular element of a friendship enriches my life.

Interests in commitment are the same. If I begin by looking at a field of commitment-objects, there will be many that I will prefer because of good-making features that many others would agree upon (for instance, that one has an interest in having a commitment to family because family will provide love and support), and others that I will prefer for more idiosyncratic reasons (for instance, I may see myself as having an interest in having a commitment to adventure sports because I love the rush of adrenaline). But if I turn from the commitments that I might acquire to the commitments that I already have, then I might see as good-making some features that I would not have seen as good-making if I were choosing them from a set to which I
had no previous attachment. Indeed, I might see as good-making features that I previously saw as down-right bad-making, and that others continue to see in this way.

Importantly, this may once again be the case whether I have intentionally held a commitment for a long period of time, or whether I have unintentionally lived in accordance with a commitment for such a long time that it has become a part of my identity. This is important for adaptive preferences. Think, for instance, of the Sufi Pirzada women discussed in Chapter Two. They may now correctly take themselves to have an interest in remaining veiled, even if they initially began to do so only because they felt that they had no choice and only gradually came to see the veiling as a part of their identity. Outsiders may think that wearing such a restrictive and uncomfortable veil could not possibly be in a person's interest, and in an agent-neutral sense they may be right. But at this point, we are not talking about the interests of "a person" – we are talking about the interests of these particular women. In a discussion of how to make prospective judgments about what one ought to do, Elizabeth Harman cautions us to remember that "as identity changes, what it is permissible to desire changes, too." Prospectively, she is right to note that this fact gives us a reason to concentrate on more agent-neutral values as we decide who to become in the future – but we can never make these judgments without simultaneously considering, retrospectively, who we already are, and what is best for me as I actually am may not be what is agent-neutrally best for persons.

If we think of interests in this way, it can explain why paternalistic interventions are justified to bolster the positive freedom of children, but not of adults. There are stringent prohibitions against actively doing harm to a person, and acting in a way that directly violates a person's interests certainly harms her. Adults' commitments, whether or not they are agent-neutrally best, are central to their interests. This is how we can explain a prohibition against paternalistically intervening in the lives of adults. Even if external parties are right to note that an adult's preference is one that she would have been unlikely to form in most circumstances, insofar as she has a history with it and her settled adult identity has developed in accordance with it, she really does have at least some objective interest in continuing to have it – namely, the interest that springs from the commitment-object having been an object of commitment that forms a part of her adult identity.

---

As we saw, adults may lack the benefits of autonomy, which make them better able to align their subjective and objective interests. Indeed, they may even have developed commitments that preclude the exercise of the kind of independence of mind that autonomy requires, or which prevent them from seriously considering options that would count as valuable for others. They maintain, however, another tool for aligning their subjective and objective interests: a privileged epistemic access to their own lives which gives them a unique insight into how their commitments have changed them. An adult may of course change her commitments, but when she does so she must bear in mind the commitments that she already has and decide in light of the settled self that she already is whether or not it is in her interest to maintain those commitments that she might have been better off not developing in the first place. This epistemic access allows her to determine in a way that others cannot whether adopting, experiencing, or seriously considering prima facie valuable alternatives would be destructive of the person who she has become, and it allows her privileged access, as well, to determine whether being that person is valuable to her. Since external actors necessarily lack this internal epistemic access, and therefore cannot know with any certainty that a person’s commitment-related interests are outweighed by her more obvious conflicting interests, the stringent prohibition on harming requires them to refrain from paternalistic action.

The lives of children, however, have not yet been narrowed. While they will also eventually have an interest in having specific commitments which will necessarily narrow and direct the scope of their future identities, as children they have an interest in being exposed to a variety of possible commitments and ways of living. Adolescence is a period in which identities are formed and meaningful commitments begin to be made for the first time.228 While adolescents will also have a level of privileged epistemic access to their own interests, their commitments will play less of a role in determining what those interests are: their histories with their commitments will as yet be short and they should be encouraged to maintain only provisional commitments while they are still young enough that it is relatively costless to revise them. At some point, with the appropriate support, they will develop the independence of mind that will allow them to begin to autonomously choose their preferences. Ensuring that they

228 Adolescence, which is socially defined, should be distinguished from puberty, which is defined in terms of biology. Since adolescence will vary from community to community, the point at which coercive attempts to expose children to new alternatives may also vary.
continue to be exposed to a broad range of prima facie valuable options at this point will help to ensure that they eventually choose the commitments that are most likely to align with their objective interests. And since even highly autonomous adolescents will be engaged in choosing and revising their commitments and identities, there exists a justification for ensuring that they develop this independence of mind to the best of their abilities, that they are exposed to a broad variety of prima facie valuable options, and that they are protected from prima facie harmful options that may have lasting effects. This will help them to avoid an adult future in which they must weigh up the costs of maintaining a commitment that they would have been better off not having developed in the first place.

In other words, the seemingly negative elements attached to a possible commitment-object count more strongly against developing a commitment in the case of a child who does not already have an identity-related interest than it does against maintaining the commitment in the case of the adult who already has it. This means that it is more important to ensure that children seriously consider options that they have not requested when they are in the process of choosing their commitments, than it is to ensure that adults, who have a history with their own commitment-objects, consider options that they have not asked for and which may violate the commitments that they already have. To use the vocabulary of freedom once again, this means that paternalistic intervention is justified to keep the futures of children, but not of adults, open, when they tend in seemingly negative directions.

This new justification can helpfully be thought of in terms of Edna Ullmann-Margalit’s vocabulary of choosing, opting, and drifting. Choosing is “choosing for a reason”, which presupposes a set of preferences which one already has. Choosing is a valuable way of selecting options, since it allows us to draw on information that we already have about ourselves to inform the courses of action and values that we will be best served by selecting. Sometimes, however, we must make “choices” that will profoundly transform who we are, and in such instances, our previous preference set can provide only limited guidance. After all, we cannot know whether we will still endorse our earlier preferences after we have been profoundly transformed in ways that we cannot fully anticipate prospectively. In such cases, we “opt” for a value or

---

229 See Chapter Six for a discussion of how the options that are to count as valuable for children should be determined.


231 Ibid., p. 157.
course of action, rather than meaningfully choosing it. Opting, however, can be distressing, since it involves giving up the person one is. Accordingly, in what ought to be an opting situation, one may “drift” — that is, when faced with a transformative decision, one may “[carry] on with the business of her life, making incremental stepwise decisions only” while “ignoring aspects of [her] decision situation, which reveal it for what it is: a first commitment leading down a core-transforming, irreversible road”. In such cases, Ullmann-Margalit notes that the chooser engages in self-deception, which shields her from the transformative nature of her choice.

This phenomenon of drifting explains one of the processes by which adaptive preferences may form. A person in limited circumstances, pressured to comply with a norm that she does not endorse, may drift into complying with that norm on so many instances that she does not realize, except in retrospect, that this norm has become a part of who she is. Whether or not she likes it, or would have liked it from her prospective pre-drifting position, she is now stuck with who she is, at least to the extent that changing would require significant upset to her adult character. Choosing or even opting would have been better for her, since these methods of selection would have allowed her to bring her previous preferences and reasons to bear on her choice, however minimally they could have informed it in the case of opting.

The new justification, then, recognizes that drifting situations do happen and turn us, for better or for worse, into the people we are today. It recognizes that the people who we have become deserve consideration, whether or not we would have been better off not having become those people in the first place. But because it recognizes that we may have been better off not developing into those people, it aims to prevent opting situations from becoming drifting situations in the first place, and to make opting situations as open-ended and choice-like as possible. It does so by encouraging children to develop the capacity for choosing meaningfully, while also exposing them to and encouraging them to try out various prima facie valuable options — even in cases in which their ultimate selections will unavoidably look more like opting situations than choosing situations.

A Dilemma

Unfortunately, while attractive, this justification leads to a dilemma. I will not attempt to solve it here, but I will at least raise it. The dilemma is this: while preventing

---

232 Ibid., p. 158.
90
children from developing adaptive preferences requires exposing them to a wide
diversity of alternatives at an early age, exposing them to these alternatives seems to
require infringing upon the negative freedom of adults in unacceptable ways. Let me
illustrate.

Children, as we saw, can fail to be exposed to a broad selection of prima facie
valuable ways of living. Public schools provide one way of exposing children to
options, but not all children attend public schools. The Amish and some 6 million
other students in the United States alone attend private schools, and another 1.5
million are home-schooled and so do not attend school at all.234 In the developing
world, where the work of so many adaptive preferences theorists is focused, the
problem is even worse. Many children’s families cannot afford to send them to school
beyond the earliest grades, and even when children can attend, the quality of
instruction and materials is often poor. But even where children are required and able
to attend public school, they may fail to be exposed to options in good time. As John
Stuart Mill reminds us, the family is the earliest, and perhaps the most important
school of moral education.235 Many children in the developed world do not attend
school before the age of 5, and if we are to believe the theorists writing on Wisconsin v.
Yoder, much of the schooling most relevant to the development of independence of
mind does not occur until after the age of 14. But by the age of 5 – and certainly by the
age of 14 – a child already has a significant history with the commitments of her
parents. And unfortunately, at the earliest stages, children are too young to make
anything but uncomprehending proto-commitments. While some of these proto-
commitments may be in a child’s objective interests, others will not be, and in these
latter cases, we should be concerned about the role that proto-commitments can have
in increasing the path dependence of their lives.236 If a child lives long enough in
accordance even with a proto-commitment, it may begin to shape her identity. In other
words, allowing a child to develop prima facie bad proto-commitments while she is still
too young to reflect on them robs her of her chance to develop her commitments
autonomously later in life. This is one horn of the dilemma.

234 See the National Center for Education Statistics (nces.ed.gov). Data is from 2010 and 2007,
respectively.
236 For the extent to which children’s early experiences shape their aspirations, see Albert Bandura, et.
Development, 72(1), p. 187–206. For the early effect of the family on the development of gender roles,
see Susan M. MacHale et al (2003) “The Family Contexts of Gender Development in Childhood and
But how can exposing children to options at this early age be done? The problem is that much early education is informal – it involves children observing adult behavior and internalizing the values and commitments that those adults model in their everyday lives. So exposing children to a range of alternative life options will mean parents (or more plausibly, community members, since each person has only one life to live) going about their lives in a broad range of ways. And if, as I argued, people’s commitments become increasingly their own through a history of engagement with them, then we should think that there is something problematic about asking adults to no longer live their lives in accordance with their own commitments and the interests that they take themselves to have simply so that the children growing up among them can be exposed to a range of options. While that range of commitments will be instrumentally valuable for the children developing their own commitments, ensuring the development of the capacity for autonomy in children cannot justify completely violating adults’ right to autonomy. Such violation will seem especially unjustified for childless adults in a community who never chose the additional responsibilities that come with parenting.

In this case, we cannot simply appeal to the harm principle to justify coercing adults into living their lives in different ways. The harm principle would allow coercion of adults to prevent them from engaging in clearly harmful practices like giving children dangerous drugs or forcing them to engage in slave labor. But in asking adults to live in a broad range of ways, we are not asking them to refrain from living in harmful ways. It would be perfectly permissible and even beneficial to expose children to a few adults espousing the values in question – it only becomes problematic when they are exposed exclusively to those values. But since the values are permissible ones in the individual case, it is arbitrary to demand that any particular adult adopt an alternative set. And insofar as particular adults have commitments to those values that contribute to their objective interests, it would also be over-demanding to do so. This is the dilemma’s second horn.

Providing an acceptable response to adaptive preferences will require finding a solution to this dilemma.

Conclusion
Although this chapter has ended with a dilemma, it has nevertheless brought to light an important consideration for theorists of adaptive preferences. As we have seen, adults may come to have genuine interests in their preferences, including their
adaptive ones, through a history of engagement with those preferences. This recognition gives us reason to take seriously those critics of adaptive preferences who charge that the concept will license unacceptable intrusions into the lives of persons judged to have adaptive preferences, while also providing reasons against such intrusion. At the same time, however, this chapter has noted that the reasons in favor of allowing adults to maintain their adaptive preferences do not count in favor of allowing children to develop adaptive preferences themselves. If children – who for the most part cannot yet make their own decisions autonomously in any case – can be put in a place to make competent and informed decisions when and if they do develop the capacity for autonomy, then this should be done. Accordingly, I offer a final consideration:

**Consideration 10: Children.** A theory of adaptive preferences must be sensitive both to the adaptive preferences that adults already have and the adaptive preferences that children might acquire. Currently held adaptive preferences should be treated differently from those that might yet form.

Before ending the chapter, however, I want to raise one final point. The discussion in this chapter, focusing as it did on literature on children written in a developed world context, reinforces the point that adaptive preferences are by no means only a developing world problem. This should come as happy news to the theorists, like those referenced in the Introduction, who think that the whole concept of adaptive preferences tends to cast those in the developed world as effective agents who know their own preferences, and those in the developing world as vulnerable dupes of culture, patriarchy, or poverty.

It may also, however, raise another worry: if the complaint in the Yoder case is that removing children from school at the age of 14 prevents them from developing critical thinking skills necessary for autonomy, does this mean that only those who have achieved a relatively high level of formal education can be autonomous? This implication would be hardly more appealing than the implication that adaptive preferences occur only in the poor.

We should, I think, reject this implication – yet it is a suggestion that we should nevertheless take seriously. We should reject the implication because there is no good reason to think that formal education is either a necessary or a sufficient condition for the development of critical thinking skills. It is not sufficient because, as we know, some schools are much worse than others. Teachers may be poor, classrooms may be overcrowded, and curriculums may be badly designed. When these conditions come
together, we should not think that critical thinking skills will be an automatic output of a high school education. Neither should we think that a high school education is necessary for these things. Developing critical thinking skills requires exercising them, and this can happen in a wide variety of contexts – they can be developed by reading widely, by having spirited discussions about values, and by problem-solving in a variety of real-world contexts. Similarly, persons can be exposed to valuable alternatives in all kinds of both formal and informal ways.

What then is the concern of the theorists writing about Yoder? I think that we can best understand their concern if we think of compulsory high school education as a way of playing the odds. For children whose critical thinking skills are not engaged or who are not exposed to many options outside of school, making a high school education compulsory will raise the chances that these skills will be acquired – especially if the right resources are put into schools, so that they will be more likely to achieve these purposes. And students who have these opportunities outside of schools will hardly be harmed by the further exposure and exercise of their skills that happens in school. Requiring a broad high school education will be a way of casting a wide net – it is at least a fairly reliable way of engaging these skills that aims to catch children who would otherwise have failed to acquire them.

The implication that we should take seriously, then, is not that school is necessary for being able to make autonomous decisions about our lives – but we should take seriously the idea that an education nevertheless significantly increases the chances that we will be able to do so. And if we think that education is good for this – namely that it improves our lives by helping us to think more clearly about them and explore options that we might not have otherwise considered – then this should by no means be surprising. Indeed, it would be much more surprising if we were to conclude that education played no role at all in increasing our ability to make our lives our own.
In this chapter, I make an initial case for thinking of adaptive preferences as autonomy deficits, and situate the sources of suspicion related to the adaptiveness of preferences within the wider set of reasons for which non-autonomous preferences might be treated as suspicious. I then begin to forestall some of the objections that might be made against such an account. Primarily, I consider and argue against the claim that seeing adaptive preferences as autonomy deficits opens persons with adaptive preferences up to coercive paternalistic intervention and accordingly fails to show them primary recognition respect.

**Autonomy, adaptiveness, and suspicion**

In the last chapter, I focused on the literature on children and autonomy because there was no literature on children and adaptive preferences. As I also noted there, however, the way in which the concept of adaptive preferences is used in the literature lines up in important ways with the values that lie at the heart of the concept of autonomy. Most importantly, both concepts make sense of the idea that persons can in fact (at least in some sense) hold preferences and (to at least some extent) endorse norms that they ought not hold or endorse. Such an idea is, of course, integral to the political project discussed in the Introduction, since that project involves applying suspicion to those socially imposed or supported preferences that persons seem to be better off not having. That political project, however, does not by itself explain which preferences persons would be better off holding, and as I argued in Chapter One, providing a perfectionist account of the preferences that persons ought to hold makes it difficult to show persons secondary recognition respect. An explanation that refers to autonomy, however, seems much more promising; by emphasizing the importance of reflective self-rule, autonomy-based accounts of adaptive preferences allow individual persons to remain the ultimate authorities on their own good and the preferences that they ought to hold, while still allowing that persons' actual preferences can be suspect because

[^237]: See the following chapter for the ways in which this claim might be complicated.
they lack autonomy. Autonomy-based accounts of adaptive preferences, then, are promising because they both show persons secondary recognition respect and further the political project associated with the concept of adaptive preferences.

If we are to adopt an autonomy-based account of adaptive preferences, however, much will clearly hang on just what we take autonomy to be. I am by no means the first person to propose an autonomy-based account of adaptive preference. Indeed, Jon Elster, who first brought the concept to prominence in the academic literature, argued that adaptive preferences should be understood as autonomy deficits. Elster does not, however, ever provide a definition of autonomy; instead, he simply lists cases that are not instances of autonomy, and understands autonomy as what is left over. In his words,

... autonomy will have to be understood as a mere residual, as what is left after we have eliminated the desires that have been shaped by one of the mechanisms on the short list for irrational preference formation.238

Since Elster does not tell us what he means by autonomy, and since accounts of autonomy differ (as we will see in the next chapter) so significantly, his account of adaptive preferences is not particularly useful.239 It helps to explain why adaptive preferences deserve suspicion insofar as autonomous preferences are usually granted a kind of normative authority, but it does not give a clear account of why precisely adaptive preferences are worse than autonomous ones.240 Nor does it offer an account of autonomy that might provide a target for adaptive preference response. Accordingly, I will not discuss Elster’s account in detail.241

239 For more on this point, see Ben Colburn (2011) “Autonomy and Adaptive Preferences” Utilitas, 23(1). Colburn also offers an autonomy-based account of adaptive preferences, but I believe that it is too narrow to be useful, since it cannot tell us what is wrong with certain kinds of so-called “character formation”.
240 Other autonomy-based accounts of adaptive preferences are similar. Serene Khader argues that Martha Nussbaum and Susan Moller Okin both offer autonomy-based accounts of adaptive preferences. I am not convinced that Khader is correct on this point, but if she is, then she is certainly right to note that “Nussbaum and Okin do not tell us much about what they mean by ‘autonomy’, so their views on what is morally problematic about [adaptive preferences] are largely implicit.” Serene J. Khader (2012) “Must Theorizing about Adaptive Preferences Deny Women’s Agency?” Journal of Applied Philosophy, 29(4), p. 306.
241 I also prefer not to discuss Elster’s account in detail because it is, as Nussbaum says, “rather narrowly focused”. On his account, a preference can only be adaptive if it involves giving up on an actually-held previous preference because of frustrating circumstances. This renders it unable to account for those adaptive preferences that result from growing up from birth in limited circumstances, and such cases are centrally important to the political project. See Martha C. Nussbaum (2000) Women and Human Development: The Capabilities Approach, Cambridge: Cambridge University Press, esp. pp. 136-139.
Let us begin our discussion of autonomy by getting a clearer idea of the reasons for which non-autonomous preferences might deserve suspicion. Like Elster, this means beginning with a negative rather than a positive account of autonomy. But this negative account will, eventually, help us to build back up to a positive one. It will also help us to separate out the kinds of suspicion that relate specifically to adaptiveness, and accordingly show how adaptive preferences fit into the broader set of non-autonomous preferences. I divide these sources of suspicion into two types: genesis and content.

**Genesis**

First, a preference might be liable to suspicion because of the way in which it was formed. Genesis-based suspicion, however, may occur in two different ways. Think of the deliberative theorists discussed in Chapter Two. Primarily, they advocate deliberation as a way of collectively ensuring that persons’ judgments and preferences are informed and carefully scrutinized. The source of suspicion here, and the shortcoming that deliberation is designed to address, is a failure to submit preferences to the kind of critical reflection that can parse those preferences that can be reasonably endorsed by a person from those that cannot.

Yet such failure can also be of two very different kinds: broad failures of capacity, or more particular failures of exercise. In the first case, a person fails to critically reflect on a preference because she lacks the ability to do so. If a person has an extremely low capacity for critical reflection, then critical reflection clearly cannot do much to improve her preferences. In the second case, a person has the capacity for critical reflection but, for some reason, fails to exercise it. Various reasons for failure include laziness and distraction. While these two kinds of failures may offer very different prospects for possible future successes of critical reflection, the immediate outcome for the preference will often be the same. If laziness, for instance, prevents a person from reflecting in any way upon a preference that she has formed, then it should be almost as suspect as the same preference held by a person who almost entirely lacks the ability to critically reflect at all. In both cases, persons are unlikely to have happened to alight upon and so prefer the option that is best for them.

Failures of critical reflection are internal failures related to genesis, but genesis is also affected by external circumstances. Whether a person critically reflects on her

---

242 Of course, these failures of critical reflection can also compound one another: a person with a fairly low level of capacity for critical reflection might, for some reason, fail to exercise it.

243 Recall from the previous chapter that I favour an informed-desire account of persons’ interests.
preferences as she develops them affects the degree of suspicion that those preferences warrant, but the circumstances under which she undertakes this reflection are also important. If a preference is developed without even an awareness of alternatives, then critical reflection will do comparatively little to determine whether it is a preference that should be maintained. A preference developed in this way, then, should be treated as much more suspect than one that is upheld in light of various alternatives. For instance, a person's preference to study mathematics will be much less suspect if she has taken some courses in other subjects and rejected them than if she has only ever tried math.

This taxonomy helps to identify two important areas of concern, but taken together, they can lead to a new cause for suspicion. Think, for instance, of a case involving a hypnotist. Imagine that I have the capacity for critical reflection, but that a hypnotist induces in me a preference to study mathematics at university which replaces my previous preference to study philosophy. Should the reason to second-guess this preference be taken to be a failure of internal critical reflection or a failure of conducive external circumstances? There are cases to be made in each direction. Clearly, the failure of critical reasoning is a result of external circumstances. If the hypnotist had not intervened, then the failure would not have occurred. The real problem with the preference, however, is not that it was developed without awareness of alternatives in the external environment, but rather that external circumstances compromised the internal process of critical reflection that would have allowed them to be seriously considered. Such combined instances of internal and external sources of suspicion should be treated separately, since different things will be required to overcome suspicion in such cases. Cases that are only failures of external circumstance will only require the introduction of new alternatives. Cases that are only failures of the exercise of capacity will require, straightforwardly, exercising the capacity. But when external circumstances have been internalized to the extent that they make a person disinclined to critically consider alternatives that become newly available, overcoming suspicion will mean not just becoming exposed to new alternatives, but also becoming willing to see them as alternatives that are genuine and that deserve critical consideration.

Clearly, all of the sources of suspicion that arise from genesis are relevant to autonomy. Failures of capacity for critical reflection, failures of circumstance, and internalization of limited circumstances all rob persons of the tools to exercise effective self-rule; the failure to exercise critical reflection simply leaves these tools unused. But
are they all relevant to adaptive preferences? They are not. The word "adaptive" itself is necessarily attached to the idea of adapting to circumstances, and the political project behind adaptive preferences is concerned particularly with preferences that warrant suspicion because they have been formed in accordance with circumstances of social injustice, inequality, or domination. Accordingly, therefore, simple failures of exercise of capacity for critical reflection are not relevant to adaptive preference, since they are not caused by circumstance. Failures of capacity might be related to social injustice or inequality, as when infants in impoverished households fail to receive the medical care, special education, or nutrition that they need to develop their capacity for critical reflection. However, although their future preferences will likely be affected by the failure to develop this capacity, the suspicion that their preferences warrant will be a result of direct causation by their circumstances, rather than moderated adaptive response to them. Accordingly, I also take deficiencies in capacity to fall outside of discussions of adaptive preferences. Both external and internalized failures of circumstances, on the other hand, will be relevant to discussions of adaptive preferences – although the former will likely be easier to counteract than the latter.

Content

Content-related reasons to be suspicious of preferences most clearly connect with perfectionist commitments about objective value. As we saw in Chapter One, it can be incredibly difficult to make uncontroversial substantive value claims of the sort that generally ground claims of the adaptiveness (or more broadly, the badness) of preferences. Once one has committed to some account of value, however, a clash between the content of that account and the content of a preference is an obvious reason to treat the preference as liable to suspicion. Indeed, this is the route that is most commonly taken by theorists who employ the kind of intuitive account of adaptive preferences discussed in the first half of the thesis. They label a preference as adaptive when it seems to be bad for a person, unworthy of a human being, flourishing-incompatible, or something else along these lines. And while one need not be a perfectionist in order to count a preference as liable to suspicion on the basis of its objective content, perfectionists will be less likely than non-perfectionists to consider the possibility that this suspicion may potentially be unwarranted.

Sources of suspicion related to content, however, may also concern subjective or personal value commitments. Here, a preference will be liable to suspicion not because its content conflicts with the content of some purportedly objective list of values, but
rather because it exhibits a conflict between a person’s own various subjective value commitments. This sort of conflict is compatible with endorsing an objective account of value, and may be cause for concern even when no preferences that contravene that objective account exist. Any plausible account of objective value will need to be both broad and pluralist, and different elements of such an account may conflict with one another. Even on an objective account of value, then, a person may hold conflicting value-commitments that would each be acceptable on their own. The fact that they conflict, however, should make them both liable to suspicion since satisfying one will necessitate frustrating the other – a state of affairs that will almost certainly make a person worse off on at least some important measure than she would be if none of her preferences needed to be frustrated. Note the way, then, in which this element of the content category relates to the genesis category: suspicion on the grounds of subjective content will often be warranted precisely on the grounds that it suggests that a person has failed to critically reflect on her preferences and bring them into line with one another.

Now, either of these categories may be relevant to autonomy. If it is the case that there are objective values that exist independently of human beings, then failing to choose in accordance with them might compromise the quality of your self-rule. I consider this possibility in the next chapter. Conflicts between subjective or individual values, on the other hand, are more obviously related to autonomy. Insofar as a person has conflicting values, she will always fail to regulate her life in accordance with some of them. Either of these categories may, as well, be relevant to adaptive preferences. Theorists like Khader and Nussbaum, as we saw in Chapter One, make objective content a central element of their definitions of adaptiveness. Subjective value conflicts, as well, may be the effect of adaptation to unjust circumstances. Imagine, for instance, that a woman develops a desire to do groundbreaking scientific research, but also develops a distrust of women as serious scientists because of the sexist norms that surround me. Subjective value conflicts, then, may be but need not necessarily be relevant to adaptive preferences.

The various sources of suspicion can be arranged in something like the following table.
Importantly, in most of these cases, the fact that a person has a reason to reconsider her preference does not necessarily mean that she has a reason to reject it. It is completely possible that she could have hit on a preference that tracks her interests without having reflected on it, or having been exposed to alternatives. And once she has considered the alternatives, she may well continue to endorse her previous preference – but insofar as she cares that her preferences reflect her interests, or that they are “authentic” or meaningfully her own, she should at least undertake this reflection in light of alternatives. The problem with an unreflective or uninformed preference is not that its object is of necessity bad. Instead, the problem is that it could be, and that she will have no way of knowing that it is without considering the alternatives available. The case of conflicting subjective preferences is somewhat different, since creating a coherent preference set will mean rejecting at least one of the preferences involved. But once again, nothing need necessarily be wrong with the object of either preference – so once again, the fact that a preference is non-autonomous, or liable to suspicion, need not be understood as a reason to reject it.

Indeed, Serene Khader has suggested to me that well-designed societies will be structured precisely so that persons fall into valuable default modes of functioning without having to reflect on the matter. I agree with this point broadly, but doubt that it will hold in more specific cases. My society may be designed so that I unthinkingly take up soccer, for instance, which will be good for me insofar as it helps me to remain fit and active. It is likely, however, that if I critically reflect, I will adopt an alternative way of keeping myself healthy and active that is better suited to me in particular. In this way, my preference remains suspect, while still being good for me.
In this respect, the liability to suspicion that comes from a failure to prefer objective value will be different. In this case, there is indeed something wrong with the object of the preference, and recognizing this kind of liability does constitute a reason to reject a preference rather than only to reconsider it. However, the controversy here seems to me to be in identifying the actual content that is supposed to be objectively valuable, not in the question of whether a person has a reason to reconsider her preference if it is for something that is of objective disvalue. When a person herself realizes that she prefers something that is bad for her, then she clearly has reason to reject her preference. The difficulty lies in convincing her that her particular preference is objectively bad for her if she does not believe that it is.

Autonomy and primary recognition respect

At this point, autonomy seems to be an attractive basis for a conceptual account of adaptive preferences, because it both succeeds in showing persons secondary recognition respect and explains the liability to suspicion associated with adaptive preferences. But before I offer a more complete account of adaptive preferences as autonomy deficits, I want to lay aside one potentially devastating objection: autonomy-based accounts of adaptive preferences may be well placed to show persons secondary recognition respect – but can they show persons primary recognition respect? That is, while it can make the space to treat persons as competent moral reasoners, can it also show sufficient appreciation for the interest that they have in directing the course of their own lives?

Now, non-autonomous people who realize that their preferences are not autonomous have at least one reason not to act on those preferences. This is the well-being related reason that I offered in the last chapter: non-autonomous preferences are less likely than autonomous preferences to line up with a person's objective interests. In the last chapter, I gave a well-being related reason to think that this reason to refrain from action does not also provide a reason for third parties to prevent persons from acting on their non-autonomous preferences: namely, because they do not have access to the way in which a person's commitments and interests interact, and might therefore undertake a course of action that would harm the person in question rather than helping her. Not everyone, however, will find well-being based arguments compelling. In the remainder of this chapter, I offer an argument against coercively

\[245\] See the next chapter for further complications on this point.

102
intervening in the lives of persons with non-autonomous preferences that is based in
the value of autonomy rather than the related and narrower value of well-being. As
Ann E. Cudd rightly notes, there are many different justifications for thinking that
actually possessing autonomy (or, as she calls it, freedom) is important and to be
worked toward,\textsuperscript{246} and the argument that I offer is meant to be compelling to anyone
who takes the possession of autonomy to be important for whatever reason.

The concern that conceiving of adaptive preferences as autonomy deficits might
open persons up to primary recognition disrespect is most clearly articulated by
Serene Khader. Citing autonomy’s use in political theory as “the capacity that exempts
people from being subject to coercion,” she argues that thinking of adaptive
preferences as autonomy deficits may have the unsavory consequence of justifying
coercive intervention into the lives of persons with adaptive preferences.\textsuperscript{247} In order to
understand Khader’s worry, it will be helpful to revisit the distinction in the last
chapter between the two ways in which the term “autonomy” can be used. The
literature on autonomy frequently draws a distinction between descriptive claims
about autonomy and normative claims about autonomy, although, unsurprisingly,
different theorists use different terms to make the distinction. Thomas Hill
distinguishes between “autonomy as a right” and “autonomy as a psychological
capacity”\textsuperscript{248}, while Joel Feinberg distinguishes between the “capacity” for autonomy
and the “right to self-determination.”\textsuperscript{249} Marina Oshana refers to “personal autonomy”
and “political autonomy”, respectively.\textsuperscript{250} Although the terms differ, however, their
meaning is essentially the same. What I call descriptive autonomy refers to a person’s
possession of various characteristics that allow her to function in characteristically
autonomous ways, while what I call normative autonomy refers to a person’s right to
be free from certain types of interference when she makes certain centrally important
decisions about her life. Khader’s worry about defining adaptive preferences in terms of
autonomy hinges on the idea that it is the descriptive claim about autonomy that
grounds the normative one – that a person who has descriptive autonomy is owed

\textsuperscript{247} Serene J. Khader (2011) Adaptive Preferences and Women’s Empowerment, Oxford: Oxford University
Press, p. 104.
esp. p. 14 and following.
normative autonomy, and that a person who lacks descriptive autonomy may be denied normative autonomy.

Khader is not the only theorist to make this connection. Indeed, other theorists writing on autonomy recognize a similar relationship between descriptive autonomy and normative autonomy, even if they do not explicitly distinguish between the two. John Christman holds that autonomy is that which "places limits on paternalistic interferences or manipulative dealings with [an agent]," and Onora O'Neill makes a similar observation in the narrower context of medical ethics. Catriona Mackenzie, meanwhile, notes that the requirement to respect autonomy in others has its base in "the presumption that autonomy confers normative authority over one's life; [that is] the authority to make decisions of practical importance to one's life." Of course, the lack of descriptive autonomy is not by itself enough to justify intervention into persons' lives – the intervention must also be for persons' own good. But when persons are not equipped to make decisions for themselves about what benefits them, the argument goes, outside parties are justified in ensuring that non-autonomous people do function in ways that benefit them. And although none of the latter set of authors mentioned discuss the problem of adaptive preferences, they would all seem likely to agree: since adaptive preferences are precisely the kinds of preferences that prevent people from recognizing their own advantage or best interests, and given this connection between descriptive and normative autonomy, construing adaptive preferences as autonomy deficits would seem to endanger the rights of individuals to make decisions about their own lives. If making a claim about a person having an adaptive preference entails making a claim about her lack of descriptive autonomy, then it also seems to suggest that she is no longer owed normative autonomy. And this will be an unpalatable consequence for anyone who cares about adaptive preferences because of their connection with inequality and marginalization: the concept would be distasteful indeed if its use served to further marginalize persons with adaptive preferences by marking them as unentitled to make important choices about their own lives.

One possible response in the broad autonomy literature to the worrisome consequences of this connection between descriptive autonomy and normative autonomy.

---

autonomy is to insist that theorists employ a thinner, less demanding account of descriptive autonomy. Marilyn Friedman, for instance, notes approvingly that “reducing the requirements for autonomy may minimize the number of persons whose autonomy goes publicly unrecognized.”254 And in the same vein, Christman argues that it is “sensitivity to pluralism and difference [that] compels us toward a maximally flexible notion of autonomy.”255 The motivation for these moves towards a less demanding conception of descriptive autonomy has at least one reason to look attractive to theorists concerned with adaptive preferences: the autonomy theorists who propose the move also want to ensure that already marginalized persons are not additionally denied the right to make decisions about their own lives.256 But such a move, as I will show in the next chapter, makes it difficult to define adaptive preferences in terms of autonomy, since such an intentionally broad account of autonomy will count almost no preferences as non-autonomous. As the bar is lowered to ensure that the preferences of more marginalized people are counted as autonomous, the concept of autonomy becomes less useful for picking out those of a person’s preferences that themselves contribute to her marginalization. In other words, an account of autonomy that counts almost no preferences as non-autonomous will also count almost no preferences as adaptive on the grounds of a lack of autonomy. To be sure, such a re-labeling would have the superficial effect of significantly lowering the number of preferences that would count as adaptive. Such a change in label, however, would not similarly change what is worrisome about adaptive preferences: that is, the component reasons for which they are liable to suspicion, and the relationship of those reasons to social injustice, inequality, and domination.

In the remainder of this chapter, then, I want to show that the possibility of understanding adaptive preferences as autonomy deficits can be saved without simultaneously denying decision-making rights to persons with adaptive preferences. In fact, doing so only requires making one common-sense observation: that autonomy is not a binary attribute. Recognizing this fact, however, has two parts, which are not equally uncontroversial. First, autonomy must be seen as a scalar rather than a binary

256 Khader recognizes this possibility in the context of adaptive preferences, but rejects it for two reasons. Khader (2011) Adaptive Preferences and Women’s Empowerment, p. 104-106. One involves what I call secondary recognition respect, and I return to this point below. The other involves the claim that a thinner account of autonomy cannot make sense of the deprivation involved in adaptive preference. I agree with the second point and discuss it further in the next chapter.
attribute, so that the extent to which autonomy is present is taken to be a matter of degree. But second, and perhaps more controversially, it must be seen as a selective attribute that attaches to individual preferences rather than whole persons, so that a person may have both high-autonomy and low-autonomy preferences at the same time. Before I can show how such a conception solves the problem at hand, however, I must show that this second and more controversial element is a plausible way of conceiving of autonomy.

Colloquially, it is common to hear talk of "the autonomous person," and nearly every author writing on the subject will use the phrase at some point in their work. It is not always clear, however, what the phrase is meant to connote. On the one hand, it can explicitly mean, as it does for Gerald Dworkin, that autonomy is a "global" characteristic of whole persons. In his account of the concept, Dworkin uses the term autonomy to refer to "a dimension of assessment that evaluates a whole way of living one's life."\(^{257}\) And given the common colloquial use of the phrase, such an understanding might seem intuitively attractive.

The term can also, however, be used more locally, and this is the use that I will argue is both more conceptually appropriate and more desirable. In his early work, Christman recommends such a usage, cautioning against seeing autonomy as an "all-or-nothing property" of a person's whole life.\(^{258}\) To see why this use might be more appropriate, consider two common and uncontroversial examples of a lack of autonomy: hypnosis and phobia. In the case of the person under hypnosis, autonomy might plausibly seem to be absent from the agent in a global sense. Theoretically, all of the person's commitments and preferences can be overruled by the hypnotist without the person hypnotized reflecting on or even being aware of the change. But while I would not want to deny that hypnosis should count as an example of absent autonomy, it is a fairly rare example — hardly anyone ever finds themselves in this position, and it should not, therefore, be expected to be a particularly instructive example for use in understanding the broader concept of autonomy.

---

\(^{257}\) Gerald Dworkin (1988) *The Theory and Practice of Autonomy*, Cambridge: Cambridge University Press, p. 16. Marina Oshana also understands autonomy as a global property of a person's life. She also, however, recognizes that the term may be used as well in the "local" sense, where it refers to the way in which a person acts in particular circumstances. Marina Oshana (2003) "How Much Should We Value Autonomy?" *Social Philosophy and Policy*, 20(2).

\(^{258}\) Christman (1991) "Autonomy and Personal History," *Canadian Journal of Philosophy*, 21(1), p. 3. Note, however, that Christman seems to reverse this position in his later work. I return to this change below.
Phobias, on the other hand, are much more common. But what should be said about the autonomy of one of the members of this large group of the phobic? Making a global claim about autonomy stick in this case is much more difficult to do. Imagine a person with a crippling fear of public speaking. Is she highly autonomous? This is an imprecise description, given the irrational fear that plays such a significant role in some areas of her life. But is she then non-autonomous? This will also be the wrong thing to say, if she is additionally highly thoughtful about, for instance, her political commitments. But it is even less appropriate to simply split the difference and say that she is moderately autonomous, since this description misrepresents her autonomy in both local areas. Rather, it is more appropriate to describe those areas separately in order to describe them accurately: in some areas this person is highly autonomous, while in others her autonomy is significantly compromised.

Let me briefly forestall one possible objection at this point. One may object that seeing autonomy as a selective feature of individual preferences rather than whole persons makes it difficult to make sense of the kind of lack of autonomy that results from conflicts between the subjective content of two different preferences. Instead, coherence seems to be a feature of whole persons. In some ways, this is correct. It makes no sense to speak of the coherence of a preference on its own, unless one wants to begin to make claims about the preference’s objective value. But where such conflicts occur, they will still be limited in most cases to particular bundles of preferences, and have little to do with the overall autonomy of a person. If I have simultaneous preferences to be an independent woman and to let men pay on dates, these preferences may conflict with one another— but I might still be a highly autonomous person in every other area of my life. If this is the case, then it is once again inappropriate to make a global claim about my autonomy that, as it were, averages my conflicting desires regarding dating together with my broader high level of autonomy. As in the previous case, there are areas in which I am highly autonomous.

---


and other areas in which I am not — but the areas in which I am not may have little to no impact on the areas in which I am.

Returning briefly to the example of the phobia, note that this example also supports the less controversial idea of the scalar nature of the autonomy of preferences. The person in the example above might be more or less autonomous in relationship to her phobia: it might affect her so badly that she cannot overcome it even in the most important instances, or it might be something which frustrates her and interferes with her quality of life, but which she can generally overcome — albeit with difficulty and anxiety. Both possibilities would count as instances of lowered autonomy, but they are clearly compromising to very different degrees.

A brief note, then, on terminology: like the other authors who write on autonomy, I will also sometimes use the phrase “autonomous person.” I want, however, to be explicit about what I take this phrase to mean. When I use the phrase, I intend it as shorthand for “a person who is highly autonomous in most of the more important areas of her life.” But in accordance with what I have just said, there are two things that this shorthand should not be taken to suggest. First, it should not be taken to suggest that I am actually referring to a global trait about the person in question. Rather, the autonomy of a person, insofar as it makes sense to talk about it at all, must supervene on the autonomy of her various local preferences. And second, the shorthand should not be taken to suggest that the “autonomous” should be contrasted with the “non-autonomous”. Rather, autonomy, as I understand it, is always a matter of degree: the more autonomous preference will always be contrasted with the less autonomous one. So when I compare the more and less autonomous person, I intend autonomy to be seen as non-binary in two different ways. First, a person may be more or less autonomous depending on the proportion and type of autonomous preferences that she has. And second, a person’s individual preferences can themselves be more or less autonomous.

Selective autonomy and primary recognition respect

To see how treating preferences as the locus of autonomy is useful for constructing an account of adaptive preferences that can show persons primary recognition respect, it will be helpful to begin by noting the consequences of seeing the entire person as the locus. In Christman’s later work, he reverses his earlier argument against seeing autonomy as an “all-or-nothing property” of persons. He writes that
One fairly rough way of understanding the idea that it is the descriptive autonomy of the whole person that matters in determining whether a person has the right to make certain important decisions about her own life (in Christman’s case collectively, with others), is to see descriptive autonomy as a broad and uniform capacity of a person. In this case, either a person has it or she doesn’t. Of course, descriptive autonomy on this account would not strictly be binary: capacities above the line would vary from person to person, as would capacities below the line. But in each case, what would matter would be the person’s relationship to that line. Either she would be over the line and so entitled to normative autonomy, or else she would be below the line and would not be so entitled.

If descriptive autonomy is seen as this kind of uniform capacity, then conceiving of adaptive preferences as autonomy deficits could indeed warrant the worry that Khader raises. If the existence of an adaptive preference or set of adaptive preferences were, so to speak, enough to push a person below the qualifying line for descriptive autonomy, then they would thereby exclude her from membership in the group of people accorded the right to make decisions about their own lives. Thinking of adaptive preferences as autonomy deficits, then, would indeed make people liable to violations of primary recognition respect – since they lacked the uniform capacity for autonomy, the whole range of centrally important decisions about the way that their lives go would no longer be theirs to make.

Attaching descriptive autonomy to preferences rather than persons would go a long way towards easing this worry. If descriptive autonomy were so attached, then recognizing the existence of an adaptive preference would no longer jeopardize a person’s right to make decisions across the board. Instead, the existence of an adaptive preference would have little to no effect on the descriptive autonomy of a person’s other preferences and there would consequently be no reason to deny her normative autonomy in any of these other areas.\textsuperscript{262} The consequences of seeing adaptive


\textsuperscript{262} Although it may, of course, be difficult to determine in practice which preferences are in fact non-autonomous, and the extent to which different decisions and areas of life are affected by some non-autonomous preference.
preferences as autonomy deficits, then, become far less severe when descriptive autonomy is no longer taken to be a global fact about persons.

This step alone, however, would not be enough to satisfy Khader. She recognizes the possibility of distinguishing between local and global attributions of descriptive autonomy, and holds that such a distinction fails to adequately protect persons from coercion:

If an autonomous agent makes a nonautonomous choice, why not coerce her in the domain of life where she has made the nonautonomous choice while respecting her capacities for choice in other domains of life?²⁶³

So to satisfy Khader, I will need to show how defining adaptive preferences in terms of descriptive autonomy can be compatible with showing primary recognition respect even in the case of persons' adaptive preferences.

And indeed, a similar issue will arise from a more sophisticated and accurate reading of Christman’s turn towards global autonomy. While Christman wrote in the quote above that it will be the autonomy of the person as such that will matter, he still recognizes that autonomy is technically a trait of preferences rather than whole persons. In saying that it is the autonomy of persons as such that will matter, he actually means not that autonomy should be thought of as a uniform fact about persons, but rather that it is the autonomy of certain kinds of preferences that matters most for attributing autonomy to an agent as such. On his actual account, it is the autonomy of the “commitments and frameworks of judgment that ground a broad range of decisions, tastes, and actions for the agent” that justify allowing that agent to engage in collective decision-making to determine the course of her own life.²⁶⁴ So Christman actually does see autonomy as an attribute of individual preferences, but sees different kinds of preferences as differentially important in persons' lives. This has interesting consequences for questions about adaptive preferences. On his account, adaptive preferences that did not relate to central commitments would not seem to justify denying a person the right to engage in collective decision-making, or to justify any other kind of withdrawal of primary recognition respect. But on his account, adaptive preferences conceived of as autonomy deficits would still seem to justify such a withdrawal of respect when they did relate to central commitments and frameworks of judgment. So to really answer the worry associated with primary recognition

respect, I need to show why thinking of adaptive preferences as autonomy deficits does
not justify a withdrawal of primary recognition respect in the case of either trivial or
deeply important, orienting preferences.

Scalar autonomy and primary recognition respect
The best way to do this, I think, is to point to the other way in which the autonomy of
preferences is non-binary: in addition to being selective, it is also scalar. Recall that
this meant that the autonomy of a preference, like a person, is a matter of degree. I will
outline my proposed account of autonomy in more detail in the next chapter, but for
now, I will focus on the way in which the autonomy of a preference depends on three
of the sources of suspicion discussed above.

The first of these is capacity for critical thought. Nearly every account of
autonomy includes a capacity condition, and while they vary from account to account,
such a condition nearly always involves the ability to critically reflect upon and either
endorse or reject one’s preferences. I also understand the capacity condition to include
the internal ability to reform rejected preferences in order to bring them into line with
those that have been endorsed.

Many accounts of autonomy, however, treat capacity as the whole of autonomy.
This, I think is a mistake. We must also recognize the other genesis factors that can
influence the autonomy of a preference. The second factor is, of course, the set of
external constraints brought to bear on the formation of a person’s preferences.
Constraints of this sort can take many forms: as noted above, if a person lacks relevant
information, then even a preference that has been critically reflected upon will be less
autonomous than one that is better informed. But similarly, the presence of
intimidation or the absence of live alternatives can act as constraints on preference
formation. Even where capacity exists, persons experiencing intimidation may be
understandably unwilling to exercise it, and where live options do not exist, a desire to
avoid frustration may act as an incentive to remain ignorant of the desired possibilities
that critical reflection might bring to light.265

Finally, actual engagement of one’s capacity is the last factor upon which the
autonomy of a person’s preferences depends. As I noted above, it is possible that a
person with a developed capacity for critical reflection and no significant constraints
on the formation of her preferences may simply be unwilling to engage that capacity,

265 This is, of course, the paradigm case of adaptive preferences.
whether out of laziness, disinterest, or for some other reason entirely. But as I also noted, willingness will often be affected by constraints. The absence of live alternatives or the presence of intimidation might provide a good reason for a person to be unwilling to engage her capacity when doing so seems likely to have negative consequences.

When the autonomy of a preference is understood to be scalar, as I have suggested that it should be, these factors jointly determine where on the autonomy scale the preference falls. Each factor must come into play to some degree, however, for the preference to count as autonomous at all. When a person lacks the capacity to reflect on her preferences, those preferences cannot count as autonomous, even if she is highly willing to reflect and experiences no external constraints. Similarly, a person experiencing no constraints and in possession of capacity can still have a non-autonomous preference if she is simply unwilling to critically reflect on it. It is harder to imagine a situation in which the constraints factor could be sufficient to count a preference non-autonomous, since even most highly-constraining circumstances would still leave room for some critical reflection given capacity and willingness. It might be possible to construct a thought experiment where, for instance, a complete absence of information could preclude autonomy given capacity and willingness, but in the real world, I am only committed to the weaker idea that constraints will always limit autonomy, even when capacity and willingness are present. Generally, however, each of these factors will come into play at some moderate level, and in these cases, an increase in any one of them will lead to an overall increase in autonomy. Similarly, a decrease in one may be compensated for by an increase in another.

What are the consequences for the connection between normative autonomy and descriptive autonomy? In the case where capacity is absent, or virtually absent, the consequences are minor. In such cases, autonomy can justifiably be denied, since capacity is a precondition of exercise, and since capacity is what allows a person to take advantage of circumstances that provide a variety of choices. Capacity, as well, comes closer to being a global fact about a person than do the other genesis-related sources of suspicion, and there are cases in which it places hard and fast constraints on the other sources. This explains why important questions concerning very young children, then, or those with extremely severe mental illness or cognitive disability that significantly

\(^{266}\) This point is controversial. Many theorists think that actual reflection should not be required for autonomy. I return to this issue in the next chapter.
diminishes global capacity, will still rightly be made by someone other than the person herself. But such a lack of capacity can only play this role when a certain condition obtains: it must be the case that nothing can be done at the moment to increase this capacity. Children’s capacity can certainly be increased, but only in the future when certain developmental milestones have been reached. 267 And in the case of some mental illnesses that reduce capacity, medication is available that can at least to some degree reverse the effects on capacity.

In the great majority of cases, however, seeing descriptive autonomy as scalar should be grounds for a significant reconsideration of the connection between descriptive autonomy and normative autonomy. This is so for two reasons. First, we should reconsider because of the importance that is placed on descriptive autonomy: if it is important enough to ground the right to lead the kind of life that a person wants to live, then it is something that should be enhanced whenever possible rather than just identified. Second, and relatedly, in most cases something can be done to increase the autonomy of a preference. Each of the three factors that determine the autonomy of a preference can be nudged by outsiders to some degree, and it will sometimes be possible to do this against an agent’s wishes. Capacity can usually be enhanced by requiring or offering education, or engaging a person in the kinds of critical discussion that exercise and thereby enhance critical reasoning skills. Constraints can be removed by providing new options and information. And willingness to actually exercise capacity may be increased through the encouragement of other agents.

But in the end, actually having an autonomous preference requires an agent to in fact engage in critical reflection on her values in circumstances relatively free of constraints, and such actual critical reflection in turn requires both capacity and the willingness to exercise it. When a person’s choices are made for her in cases where her preferences are non-autonomous – as they would be if normative autonomy was grounded in descriptive autonomy at both high and low levels of autonomy – then two distasteful consequences will be likely to follow. First, such a scenario is likely to undermine one’s motivation to enhance the autonomy of her own preferences where capacity is present. If a person’s preferences play no role in the way that her life will go, then she has one less reason to seriously consider them. In an important sense, they do not matter – her important decisions will be made for her, with no reference to

267 Recall, however, my argument in the last chapter: once capacity has been developed in children, there remain other justifiable reasons to compromise children’s normative autonomy, if doing so is required to expose them to valuable options while their identities are still very much in the process of forming.
them, so why should she care if they reflect her actual desires or not? But second, and perhaps more disturbingly, this imposition of external decision-making will rob the person involved of an important chance to strengthen the capacity for critical reflection in instances where that capacity is weak. The capacity for critical reflection depends in large part upon the ability to creatively identify available possibilities, the ability to judge the likelihood of certain consequences that may follow, and the ability to weigh the goodness and badness of those consequences against one another. But because the world is complicated and plausible conclusions can be elusive, each of these abilities requires exercise, trial, and error, in order to grow stronger. When a person is denied the chance to make her own choices and experience the consequences that come with them, she loses a chance to exercise all three of these abilities. The loss of such chances is especially serious because of the global nature of capacity. Although capacity is applied and exercised in specific cases, gains achieved in specific cases can nevertheless be translated and applied in other areas of a person’s life. It is for this reason that the absence of music and art education for children is so tragic. Beyond the intrinsic value of art and music, children who study them learn more broadly to think in ways that are new, creative, and provide different perspectives for considering the world.

These consequences provide a reason to reconsider the relationship between descriptive autonomy and normative autonomy. If autonomy matters enough that the right to direct one’s own life is tied to it, then there is strong reason to enhance it wherever that enhancement is possible. So instead of seeing descriptive autonomy as a precondition for the attribution of normative autonomy (at least beyond the lowest levels), attribution of normative autonomy should be seem as a prerequisite for the enhancement of descriptive autonomy. Certainly, persons with normative autonomy might not always exercise it in a way that enhances their descriptive autonomy – but they will be still less likely to undertake behavior that enhances their descriptive autonomy if the meaningful opportunities for exercise that come with normative autonomy are denied to them.

Hugh LaFollette argues, along similar lines, that parents must extend greater autonomy to their children as the latter grow up in order to strengthen their capacity for it. He does not, however, note that adults may similarly require the chance to strengthen their capacities. Hugh LaFollette (1998) “Circumscribed Autonomy: Children, Care, and Custody” in J. Bartowiack and U. Narayan (eds.), Having and Raising Children, State College, PA: Penn State Press.

Although, of course, a lack of willingness, the internalization of limited circumstances, or other factors can make it harder to apply these gains in specific areas.
Seeing autonomy as non-binary in both senses, then, should answer Khader’s objection to understanding adaptive preferences as autonomy deficits. Seeing them as autonomy deficits of the kind that I have suggested encourages seeing them as something that persons should maintain the possibility of reforming into autonomous preferences, rather than as preferences that persons should be prevented from acting upon. And seeing autonomy as doubly non-binary also explains why persons should nevertheless be granted the right to act or deliberate even when some of their most important, organizing preferences and commitments are non-autonomous. If it is important to enhance autonomy, then it will be most important to do so in the context of the commitments that have the greatest role to play in persons’ lives.

Granting normative autonomy to persons, however, is perfectly compatible with action that seeks to enhance the factors that contribute to their descriptive autonomy. Indeed, understanding descriptive autonomy as the function of exposure to alternatives and actual exercise of critical thought as well as capacity for critical thought provides a positive case for enhancing these other areas, at least insofar as descriptive autonomy is taken to be valuable. Seeing adaptive preferences as autonomy deficits, then, is doubly useful: on the one hand, when understood correctly, it offers a justification for protecting persons with adaptive preferences from coercive interventions into their lives; on the other, it provides a justification for the kinds of non-coercive interventions that can help persons to better exercise their autonomy. That is, it provides justifications for precisely those consciousness-raising and more broadly educational programs that adaptive preference and development theorists endorse, and for the provision of new social options that can give people real opportunities for choosing in ways that challenge social norms. So while seeing adaptive preferences as autonomy deficits allows adaptive preference theorists to extend primary recognition respect to persons with adaptive preferences, it also makes it possible to recognize and help persons themselves respond to what is suspect about those preferences.

Conclusion
This chapter has primarily addressed the conceptual element of my theory of adaptive preferences. The bulk of the conceptual account, however, will come in the next chapter, when I flesh out the account of autonomy that I employ and situate my account of adaptive preferences as a subset of all non-autonomous preferences.
Nevertheless, even at this early point, several of the considerations identified in the first half of the thesis have already been addressed. Primarily, this chapter has helped to inform Consideration 4:

**Consideration 4: Interference criteria.** An account of adaptive preferences should not warrant undue interference into the lives of those with adaptive preferences; any justifiable interference must be consistent with showing individuals with adaptive preferences primary recognition respect, and an account of adaptive preferences should be developed in such a way as to ensure that it is compatible with showing this respect.

In this chapter, I have argued that understanding autonomy as both scalar and selective enables an account of adaptive preferences as autonomy deficits to provide justification for allowing persons to make uncoerced decisions even on the basis of their adaptive preferences. This enables the account to show appropriate consideration for persons as moral agents for whom the ability to live a life in accordance with one's own convictions has great value, as primary recognition respect requires.

Seeing adaptive preferences as autonomy deficits, however, has also provided a justification for non-coercive intervention aimed at increasing the autonomy of persons' adaptive preferences. Accordingly, this chapter has also furthered the political project and addressed Consideration 8:

**Consideration 8: Facilitation of Relevant Skills.** Where possible, a theory of adaptive preferences should recommend responses to adaptive preferences that strengthen the capacities of agents beyond the situation at hand.

I argued that seeing autonomy as doubly scalar recommends allowing persons to act on their adaptive and otherwise non-autonomous preferences precisely in order to leave open the possibility that autonomy will be strengthened through exercise. And since the capacity factor of autonomy is a global feature of persons, its exercise in the case of adaptive preferences will contribute to the agent's strengthened capacity for autonomy across the board.

And finally, this chapter also solves one of the problems associated with secondary recognition respect, which requires recognizing individuals as moral agents who can competently arrive at a conception of the good. As I noted in Chapter One, one of the most important objections raised against the concept of adaptive preferences relates to the worry that the concept will treat persons with adaptive preferences as less competent moral reasoners than those who do not have adaptive preferences. And indeed, Khader raises this worry with specific reference to the possibility of conceiving
of adaptive preferences as autonomy deficits. She writes that understanding adaptive preferences as those preferences characterized by a lack of autonomy "suggests that people with [adaptive preferences] have had their capacities to make choices impaired ... Since autonomy is a feature of subjects, defining [adaptive preferences] as substantive autonomy deficits encourages condescension toward people with [adaptive preferences]."\(^{270}\) This is not, I think, the most important problem associated with secondary recognition respect, and I will spend the next chapter addressing what I take to be a more important one. Khader's worry about secondary recognition respect, however, can be dispensed with much more quickly, since it should be clear from what I have argued in this chapter that it is unfounded. Once Khader's assumption – namely that "autonomy is a feature of subjects" – falls, so too does the worry that seeing adaptive preferences as autonomy deficits will recommend treating persons with such preferences in a condescending manner. By recognizing autonomy as both selective and scalar, the account that I have offered recommends against seeing persons as falling either into the camp of the autonomous or the camp of the non-autonomous. Instead, it recognizes autonomy deficits \textit{wherever} they occur, both in those who would be seen as highly autonomous and those who would be seen as barely autonomous on a global account of autonomy. My account thereby offers a picture of autonomy that is both more realistic and less divisive than one that separates people into two binary groups. Instead, we all lack autonomy in some areas, and we are all highly autonomous in others. And since my account recognizes that autonomy is as much a function of circumstance as personal will and capacity, it makes it possible to identify the circumstances that contribute to lowered autonomy (and so to adaptive preferences) without suggesting that a person who lacks autonomy must also be an incompetent moral reasoner.

Chapter Five: Secondary Recognition Respect in a Conceptual Account of Adaptive Preference

In the last chapter, my primary aim was to show that defining adaptive preferences as autonomy deficits was consistent with showing persons primary recognition respect: that is, that doing so need not license coercive intervention into their lives. My argument in the last chapter, however, appealed only to a very thinly described account of autonomy. In this chapter, I provide a more robust account of autonomy with two purposes. First, providing such an account allows me to position adaptive preferences more specifically as a subset of all non-autonomous preferences, and to explain the relationship between these two sets. And second, providing a more detailed account of autonomy allows me to further explore the extent to which an autonomy-based account of adaptive preferences can show persons secondary recognition respect.

Content-neutral autonomy, substantive autonomy, and secondary recognition respect

One of the primary aims of this chapter is to develop a more fleshed-out account of autonomy. Given my broader project in this thesis, one of the primary desiderata for the account of autonomy that I develop here will be its ability to account for and structure the concept of adaptive preferences. As we will see, the various accounts of autonomy offered in the literature are very different from one another, and some do a much better job than others of accounting for adaptive preferences. I am open to the possibility that other accounts of autonomy may be preferable for other purposes, although I do not pursue this question here. I am skeptical, however, that it would be possible to develop an acceptable account of autonomy that was fully inconsistent with or completely unable to explain the possibility of adaptive preferences. As I argued in the last chapter, self-rule is undermined by the undue influence of external circumstances, and an account of autonomy that does not sufficiently explain the inhibiting effects of such circumstances will accordingly be only weakly connected to the spirit that motivates the broad use of the concept of autonomy.

Although my concern is ultimately with adaptive preferences, I begin this chapter by focusing broadly on the concept of autonomy. In the second half of the
chapter, I will offer a definition of adaptive preferences as autonomy deficits. But once again, as I noted in the last chapter, whether an account of adaptive preferences defined as autonomy deficits can accord with the considerations developed in the first half of the thesis will depend upon the features of the specific conception of autonomy in terms of which adaptive preferences are defined.

In the last chapter, I briefly argued that defining adaptive preferences as autonomy deficits was attractive because doing so treated persons as the ultimate authorities on their own best interests, thereby showing them secondary recognition respect. I then focused on showing that conceiving of adaptive preferences as autonomy deficits accorded with the considerations in the first half of the thesis by showing that persons with limited autonomy are still owed primary recognition respect. As we will see in this chapter, however, the claim about secondary recognition respect is somewhat more complicated than it first appeared – whether an autonomy-based account of adaptive preferences can show persons secondary recognition respect is one of the questions that depends heavily upon the way in which autonomy is understood.

Recall my initial discussion of secondary recognition respect in Chapter One. There I argued that making claims about substantive value on behalf of the state meant treating persons who reject those claims as incompetent moral agents. That is, a substantive perfectionist position will always treat a commitment that conflicts with the substance of its list as fundamentally mistaken in its aim at the good, no matter how well thought through or freely endorsed that commitment might be. Such an account is problematic because its application forces those to whom it is applied to choose between their informed and considered preferences on the one hand, and their desire to be seen as competent moral agents by their peers on the other.

The issues of substantive content and secondary recognition respect also get to the heart of a parallel debate in the literature on autonomy over whether or not substantive content should have a role to play in determining when autonomy obtains. Accounts of autonomy are of two main types: substantive, and content-neutral. Content-neutral accounts allow the possibility that any preference can be

---

271 Recall that I take the autonomy of a person to supervene on the autonomy of her preferences, although I continue to use the term "autonomous person" as useful shorthand. Although the theorists discussed in this chapter take autonomy to have different loci, including both the person and the preference, little hangs on autonomy's locus for the purpose of this chapter.

272 In the literature, content-neutral accounts are also called "procedural". I prefer the term "content-neutral" because it more clearly establishes what is at stake between these accounts and substantive ones.
autonomous, no matter what its content. This is because content-neutral accounts of autonomy are concerned with the procedure by which a preference is formed or endorsed, or the relationship in which it stands to a person’s other preferences, rather than the content of the preference that emerges from that procedure. In the most common accounts, often called “coherentist” accounts, a preference is autonomous if it can be found, upon critical reflection, to cohere with the broader set of commitments that makes up the person’s “true”, “authentic”, or more settled self. In other cases, generally referred to as “hierarchical accounts”, a person’s “true” or “authentic” self is associated specifically with her higher order preferences, and self-reflection can reveal whether or not her first-order preferences accord with her “real” values. In all cases, however, a person’s autonomy depends upon her preferences being able to survive the procedure of critical reflection that would serve to establish that they form a coherent part of her settled and considered identity.

Importantly, such accounts do not require a person to consider autonomy valuable in order to count as autonomous. It is the fact that her preference could survive a certain procedure, and not the additional fact that she treats the reflection or self-rule embodied in that procedure as a valuable end, that makes a preference autonomous. Accordingly, content-neutral accounts of autonomy allow those with commitments to obedience or tradition to autonomously deny the very value of self-rule itself (although such denials may nevertheless lead to diminished autonomy in the future). For instance, a woman may autonomously prefer to allow her husband to fully determine their use and method of birth control as long as doing so is consistent with her other deeply held (or higher order) values and preferences.

Substantive accounts of autonomy demand more of preferences. According to paradigmatic substantive accounts, in order for a preference to be autonomous, there must be some central element of its content that “accords” with the value of

---


autonomy itself, or, at least, [does not] undermine that value." 276 That is, on a substantive account, unlike on a content neutral one, a person cannot autonomously judge that her own autonomy is worthless. (As we will see, this paradigm substantive requirement can be replaced by others.) The substantive requirement, however, adds to, rather than replaces, content-neutral requirements for autonomy. On any account a preference must meet some set of content-neutral requirements, but a substantive account may nevertheless disqualify such a preference from counting as autonomous if its content is not of the right kind. So a substantive account may disqualify a woman’s preferences regarding birth control either because she prefers that her husband makes all choices or because her preferences were arbitrary or incompatible with her other values.

Content-neutral accounts clearly seem better equipped than substantive ones to show secondary recognition respect to a diverse range of persons. On them, there are no limits on what a person might competently prefer. Even a preference that seems to harm a person by decreasing her independence must be recognized as autonomous on this account if it has gone through the appropriate procedure. This means that it is entirely possible for unpopular, illiberal, or self-limiting comprehensive doctrines to be seen as autonomously adopted, and consequently, for their bearers to be seen as having competently arrived at them. The only limits to autonomy, on this account, relate to the way in which a person can autonomously prefer a thing. She can prefer the apparently harmful preference in a reflective, thoughtful, and committed way, or she can prefer it blithely, without considered reason. Only in the latter case will her preference mark her out as an incompetent moral reasoner. In the former case, she can be recognized as a competent, autonomous agent, no matter how much external parties might take issue with the content of her preferences.

Substantive accounts of autonomy, on the other hand, make it much harder to show secondary recognition respect for holders of controversial, self-limiting, or self-harming doctrines. These accounts use a pre-determined set of substantive criteria to separate sanctioned content from unsanctioned content, rather than relying upon the considered convictions of the persons whose preferences are being evaluated. Such convictions may carry some minimal weight in favor of a preference, but its status as an autonomous preference will always ultimately be determined instead on the basis of the pre-selected substantive criteria. This should be unacceptable to anyone who cares


122
about the interest that persons have in being recognized as competent moral reasoners. Showing persons secondary recognition respect means precisely giving them credit for the extent to which they have engaged in such reflection and consideration – in short, for the degree to which they have engaged in competent moral reasoning. But a substantive account of autonomy effectively voids this credit on the basis of controversial stipulated content that some people will be unable to thoughtfully endorse.

Given the importance that I have argued should be placed on secondary recognition respect, there might seem to be significant reason for me to advocate a content-neutral account of autonomy as an appropriate ground for my conceptual account of adaptive preferences. After all, a substantive account of autonomy seems to replicate the problems that arose from a perfectionist account of adaptive preferences: controversial value commitments are built into the theory that from the outset bar holders of certain unpopular doctrines from having their doctrines recognized as legitimate by their fellows.

The problem, however, with grounding an account of adaptive preferences in a content-neutral account of autonomy, is that doing so makes it difficult to count the paradigm case of adaptive preference as non-autonomous at all. The paradigm case of adaptive preference involves a person who internalizes her limited circumstances so completely that she takes them to be appropriate not only at the first order, but at higher orders as well. Content-neutral autonomy theorists, however, do not generally take such cases to be problematic. Marilyn Friedman holds that autonomy has to do with how I understand my identity in the moment – not with how I came to adopt it. According to her "[i]t does not matter whether someone’s concern is itself the product of her socialization or otherwise the result of circumstances over which she has no control." 277 If a concern does in fact matter to a person in the moment, then acting upon it will count her as autonomous no matter how it was formed. Harry Frankfurt similarly writes that the cause of a person’s values and identifications is irrelevant to her autonomy. 278

Taking such a position renders autonomy unable to account for adaptive preferences, since it means ignoring (or perhaps even endorsing) the possibility that circumstances of deprivation might be internalized so completely that they lead to

277 Ibid., p. 11.
active endorsement of preferences. Although I will not pursue the point further, let me note that I am also skeptical that such a position leads to an attractive account of autonomy more broadly. Taken to the extreme, such a position allows us to count as autonomous a preference implanted in a person by a hypnotist or a program of brainwashing if that hypnotist or brainwasher is only skillful enough to implant it at the higher level. This is an unacceptable consequence of an account of autonomy, since the idea of self-rule that grounds the concept is dissolved if external arbitrary implantation is the only or overwhelming fact that leads to endorsement of a preference.

Neither paradigm substantive accounts nor paradigm content-neutral accounts of autonomy, then, are promising grounds for an account of adaptive preferences. While substantive accounts exclude too many preferences that cannot be excluded given the fact of reasonable pluralism, content-neutral accounts are overly broad and fail to include paradigmatic instances of adaptive preferences. Nevertheless, I will argue, each kind of account is motivated by a concern that adaptive preference theorists should take seriously.

As we saw in the last chapter, content-neutral theorists champion content-neutral accounts at least in part because they want to ensure that the concept of autonomy can be extended to those persons who have so often been denied it. In John Christman's words, it is “sensitivity to pluralism and difference [that] compels us toward a maximally flexible notion of autonomy”; without such flexibility, it is precisely those in underprivileged and marginalized groups who will be denied the moniker. I take on board the motivation to ensure that members of marginalized groups are treated respectfully, as competent moral reasoners taken to be capable of directing the course of their own lives through a program of meaningful self-rule.

The motivation behind substantive accounts of autonomy, however, must also be taken seriously. Despite the shortcomings of such accounts, feminist theorists who endorse substantive accounts of autonomy are often motivated to adopt these accounts precisely out of a concern for the possibility of adaptive preferences. Natalie Stoljar, a strong advocate for substantive accounts of autonomy, worries that contenting ourselves with a content-neutral conception of autonomy will force us to count as

Note that the use of hypnotism to stop smoking, for instance, is different. In such a case, an outside force brings lower-order preferences into line with higher-order preferences that have already been thoughtfully endorsed. If hypnotism could be used to restore the forgotten preferences of those with brain damage, for instance, this might also be consistent with autonomy, insofar as the preferences restored had been thoughtfully considered and endorsed before they were forgotten.

autonomous – and so to legitimize – the choices of women raised in oppressive systems who come to endorse those systems. Anyone who takes feminist concerns seriously, she writes, should demand a stronger conception of autonomy that allows us to criticize this kind of adaptive preference. Similarly, Paul Benson defends a substantive account of autonomy in his early work because he holds that an account of autonomy must exclude preferences formed by oppressive socialization, but recognizes that “the general means by which oppressive socialization operates are no different than those through which benign socialization takes effect.” Substantive theorists agree, then, that content-neutral accounts alone cannot guard against the adoption by marginalized persons of the norms that maintain their marginalization. Caring as I do about adaptive preferences, I take this motivation on board as well: an account of autonomy should give us the tools to criticize the marginalizing norms that people themselves actually endorse.

An intermediate proposal: weak substantive accounts

Both proponents of content-neutral and substantive accounts of autonomy have put forward what I think are promising variations on the standard accounts. On the substantive side, some theorists have proposed “weak” substantive theories. These theories will perhaps be easiest to understand if they are thought of as “supplements” to content-neutral accounts of autonomy. Like content-neutral theorists, weak substantive theorists think that it is problematic to adopt an account of autonomy that automatically excludes the “traditional” ways of life, lived by many, that involve commitments that might conflict with the taxing demands made by “strong” substantive accounts. But like strong substantive autonomy theorists, they recognize that requirements of critical reflection and higher-order endorsement alone will not be enough to pick out many cases of endorsed oppressive situations that intuitively seem to be paradigm cases of limited or absent autonomy. So weak substantive theorists adopt a substantive requirement with a different locus: rather

---

284 Note that I do not claim that all, or even most “traditional” ways of life do include commitments that conflict with substantive accounts of autonomy. I recognize that all cultures are internally diverse and open to change. Nevertheless, it seems uncontroversial to say that at least some of the people living “traditional” lives do have such commitments, and the existence of this smaller group is enough of a reason to expand our definition of autonomy so that they can be included.
than placing a substantive requirement on the content of a person’s preferences, they place a requirement on the over-arching way in which that person regards herself.

A variety of different weak substantive accounts are available in the literature, and their proposals of course vary. The most common, however, require an agent to have a sense of self-worth in order to count as autonomous. Self-worth can have many dimensions. It can include, for instance, self-trust, which involves trusting one’s own convictions and judgments and seeing oneself as entitled to answer to others for them; self-esteem, which involves seeing one’s life and undertakings as meaningful and worthwhile; and self-respect, which involves seeing oneself as the moral equal of others. These related ideas have a clear unifying theme: they all contribute to a person’s sense of entitlement and internal ability to move beyond the circumstances that she endorses, should she so choose. It is this set of abilities and attitudes that weak substantive theorists think suffice to do the work of ruling out the problematic cases that content-neutral accounts include as instances of autonomy, while ruling in the traditional or obedient ways of life that people might legitimately prefer. When a person feels entitled and confident in her abilities in these ways, weak substantive theorists hold that she can be counted as autonomous even when she prefers things that do not further her own autonomy.

I wholeheartedly agree that self-worth is an important concept for those who care about the possibility of harmful norms being internalized. If a person has it, then there is much more reason for outsiders to trust her judgment about her good than there is in the case of a person who has internalized the view that she deserves no

---

285 However, the requirements are sometimes taken to be broader. Sigurdur Kristinsson’s account, for instance, requires persons to respond to reasons and to act according to their own “nomos”, which “seems to require certain qualities of character, sometimes strength of will, sometimes reflectiveness and deliberation, sometimes a certain self-acceptance.” I leave this account aside because it seems that the first and third qualities of character can be subsumed under weak substantive accounts that privilege self-trust, while the second amounts to a content-neutral condition for autonomy. Sirgudur Kristinsson (2000) “The Limits of Neutrality: Toward a Weakly Substantive Account of Autonomy” Canadian Journal of Philosophy, 30(2), p. 284.


287 Catriona Mackenzie talks about self-esteem in this sense, and indeed equates it with self-worth. Catriona Mackenzie (2008) “Relational Autonomy, Normative Authority, and Perfectionism” Journal of Social Philosophy, 39(4). Trudy Govier takes self-trust to be a precondition for self-esteem, although requires only the former for autonomy. The latter, however, she takes to be important in its own right.

288 Govier also holds that self-trust is a necessary precondition for self-respect, but requires only the former for autonomy. Although Benson does not explicitly use the term “self-respect”, it seems to be suggested by his account. Mackenzie endorses the concept more explicitly.
better than she gets. However, such a constraint should be seen as neither necessary nor (along with the procedural elements of a weak substantive account) sufficient for autonomy.

Such constraints cannot be sufficient for autonomy, because they do not ensure that a person has ever considered or been exposed to circumstances better than her own that might be worth choosing. Consider the example from Martha Nussbaum, discussed in Chapter One, in which Indian widows ranked themselves as being in acceptable health, despite the fact that they were in poor health according to a broad set of objective measures. These women might take their own projects to be important, and see themselves as moral equals who are entitled to make their own decisions about their circumstances. Accordingly, weak substantive accounts would count them as autonomous in relationship to their health-related preferences289 – even though poor health was all that they had ever experienced! This conclusion is unacceptable: the preferences of the women involved are the products of circumstances, with little to no input from the women themselves. The fact that persons would feel entitled to have other preferences if they had other genuine alternatives to consider cannot substitute for their actually possessing – or at least considering – those alternatives.

Neither is self-worth necessary for autonomy. Consider two examples. In the first, a man in a poor family routinely drains the common bank account to spend money on a series of mistresses. It may be appropriate for this man to have a low opinion of himself as partner and to correspondingly see himself as less than a moral equal in this area and as unentitled to make demands on his presumably more virtuous wife. Seeing himself in this way, however, should not count against the autonomy of his considered preference to use protection when he has sex with his wife in order to protect her sexual health. Enacting his preferences in this way is still clearly an instance of self-rule, if it is considered and supported by reasons. In the second example, a woman is approached by a micro-finance organization with the offer to grant a loan either to her or to her husband. Imagine that she has internalized the idea that she cannot oversee financial projects, and that her husband is the appropriate person to hold the loan. Although her self-worth in this area is clearly low, we should still count her as autonomous in relationship to a preference to hold the loan herself, even if it sits alongside the fear that she does not deserve the loan or will not know how to use it.

289 Recall that I use the term "preferences" to refer generically to beliefs, values, and preferences. This is a case in which it seems more appropriate to talk of beliefs than preferences, since the women do not seem so much to prefer their state of health as to believe that it is adequate.
well. This woman exercises self-rule in her choice, and the fact that she does so with difficulty and despite significant obstacles should – if anything – make that exercise more exceptional and more worthy of recognition.\textsuperscript{290}

Clearly, self-worth is a valuable thing. Insofar as adaptive preferences and a lack of self-worth go hand in hand,\textsuperscript{291} it is perhaps especially valuable to those persons who have adaptive preferences. I think, then, that theorists and practitioners who care about adaptive preferences have independent reason to undertake research and implement programs that help marginalized populations to increase their self-worth when it is low. At the conceptual level, however, the concept of adaptive preferences itself should not be grounded in an account of autonomy that builds these constraints of self-worth into itself. Nevertheless, the account of autonomy that I will propose will build in a related requirement: for preferences to count as autonomous, it must be the case that the holders of those preferences see themselves as being in a position to choose certain alternatives. But rather than seeing \textit{themselves} in a certain way, they must see their \textit{relationship to opportunities} in a certain way.

\textbf{Another intermediate proposal: historical accounts}

On the content-neutral side, John Christman has championed a historical account of autonomy. His account shares the intuition that desirable accounts of autonomy must be alert to problematic socialization, but proposes a procedural solution to the problem.

In the earliest version of his historical account, Christman argues that the best way for an account of autonomy to exclude oppressive socialization is to attend not to the preference that a person actually has, but to the process by which it was formed.\textsuperscript{292} On his early account, a preference is autonomous if a person did not resist the way that it was formed, or would not have resisted it had she attended to it. This condition is meant to rule a preference non-autonomous if a person reflects upon it and realizes that it was formed through a process that she does not endorse, such as brainwashing. The condition, however, is incomplete, because it cannot rule out cases in which the problematic process of socialization affected a person’s higher order preferences as

\textsuperscript{290} Once again, it is important to see autonomy as a feature of preferences rather than persons. If we do not, then we cannot account for a person who struggles and heroically achieves autonomy in only some areas.


well. In this case, she would be unlikely, in later life, to repudiate that process. To rule out such cases, Christman stipulates that a person failing to resist the development of a preference at either the higher or the lower level will only count as autonomous if her failure to resist did not take place under the influence of factors that inhibit self-reflection. Such factors include being in an uncontrollable rage or under the influence of heavy doses of some drugs, lacking minimal education, and lacking exposure to alternatives.

In his most recent work, the role that a preference's development plays in its autonomy is less direct. Now autonomy only requires that a person would not be alienated from her preference were she to critically reflect on it in light of the process by which it was formed. In this case, the formation of the preference might ultimately count against the person holding it – but it also may not. The child who learns to look after herself because her only parent is an alcoholic will almost certainly disapprove of the way in which her self-reliance developed, but might still value her self-reliance itself. It is for this kind of reason that Christman modifies his later account. He writes that a "person's attitude toward the processes that spawn her values is separate from, and less relevant to autonomy than, the person's attitude towards the value itself." I endorse this change. Requiring endorsement in light of reflection on development makes sense because many of us will be inclined to doubt our preferences if they were formed in a way that we cannot endorse. But if, upon appropriate reflection, we find independent reason for endorsement that outweighs our reason for doubt, then the requirement has done its work. The content of the now-endorsed preference should not be held hostage to the way in which it was formed. The difficulty lies in determining what kind of reflection counts as appropriate.

While I endorse Christman's change in requirement, I do not endorse his concurrent weakening of that requirement. The reflection on development that he requires may now be purely hypothetical: he explicitly states that autonomy "does not require that any actual reflection take place on the characteristic in question." His primary reason for making the reflection criterion hypothetical is that he thinks that many elements of our personalities were never chosen or explicitly reflected upon by

---

296 Ibid., p. 142.
297 Ibid., p. 145.
us, but can nevertheless “form an uncontroversial and authentic structure that defines
us and relative to which we can be autonomous if we are not alienated from them over
time.” I suspect that Christman wants to allow the possibility that unreflective
preferences can be autonomous because of his contention, raised in the last chapter,
that only autonomous preferences are exempt from paternalistic attempts at change.
But as I showed, such a contention is unwarranted: a proper valuation of autonomy
recommends allowing persons to act on their preferences even when they are non-
autonomous.

Furthermore, there are compelling reasons to think that autonomy in the form of
self-rule requires actual critical reflection. Imagine that there are two cafes in my
neighborhood: Mad Hatters and the Red Cafe. Neither one is more convenient than the
other, or better priced, and I have no reason to think that the quality of one is any
better than the other. Nevertheless, as it happens, I have only ever eaten at the Red
Cafe. While continuing to eat at the Red Cafe may in fact be the course of action that I
would endorse taking upon informed consideration of the alternative, it is clearly not
the case that my preference to do so is autonomous. At best, my uninformed preference
would be a happy accident completely unrelated to self-rule. It is the fact that a
preference does survive critical reflection, rather than the fact that it could, that makes
it a real instance of self-rule rather than a happy accident.

Actual reflection is also a superior requirement to hypothetical reflection because
it helps to eliminate conflict between a person’s preferences. In the last chapter, I
discussed conflict between a person’s preferences as one of the sources of suspicion
associated with non-autonomous preferences. Although I rejected treating the entire
person as the locus of autonomy, on the grounds that it prevented us from recognizing
the obvious instances in which levels of intrapersonal autonomy differed, I
nevertheless recognized that conflicts between a person’s preferences compromised her
ability to act on some autonomous preferences without violating others. This
compromises self-rule. Requiring actual reflection in light of one’s other preferences,
then, not only helps a person to determine whether her unreflective preference was a
happy accident or an unfortunate mistake – it also helps her to bring her preferences
into line with one another, or at least determine what level of internal conflict she is
willing to put up with. This enhances her capacity for exercising self-rule across the
board.

Ibid.

130
The cafe example above, however, once again raises the issue of what kind of reflection should count as sufficient and appropriate for autonomy. In particular, what *types* of experiences must a person be exposed to in order for her judgment about her past to count as authoritative? My preference about the Red Cafe only becomes autonomous when I actually consider it against Mad Hatters. But what if, in addition to Mad Hatters, which is in many ways like the Red Cafe, there were also another alternative which was very different? Would autonomy require me to consider the third alternative? Would it matter what the other qualities of that alternative were?

Christman recognizes that isolated instances of critical reflection in compromising circumstances can sometimes fail to reflect persons' settled personalities and dispositions. When this happens, he holds, these isolated instances fail to confer the "agential authority" that autonomy requires. He argues, then, that autonomy requires not just critical reflection, but what he calls *sustained* critical reflection: that is, "critical self-reflection repeated in a variety of contexts with similar evaluative results." 299 If a person consistently reaches the same positive judgment about one of her own preferences, Christman holds, this is sufficient to ensure that it is indeed hers and is therefore autonomous. So in my cafe example, it will indeed matter for my autonomy whether I try Mad Hatters and other alternatives.

But do the other qualities of the various alternatives matter? I argue that they do. It is easy to imagine a person who would reconsider or reject a preference when she considered it in light of a better option, but would maintain it across a broad selection of possibilities no better than the one that her original preference picked out. For instance, a woman in the 1950s in the United States considering a variety of secretarial and administrative jobs might maintain a considered preference to be a secretary, but might reject that preference when she sees other women taking on managerial roles. In such a case, reflection across a variety of similar circumstances would do little to ensure that the preference in question was not a mere product of circumstance. To ensure this would require sustained critical reflection across a variety of *better* circumstances.

Does Christman endorse such a requirement? It is difficult to tell, but I think not. He does not broach the subject in his discussion of sustained critical reflection, but elsewhere he writes that while open options are required for autonomy, they are required "only as a function of a person's particular commitments, not as objectively
given or philosophically specified for all persons as such." The problem with such a position, of course, is that it is compatible with the example in the above paragraph: a person might develop some commitment as a function of limited circumstances and maintain that commitment across a variety of circumstances that do not challenge that commitment – yet abandon the commitment when she considers it in light of circumstances that embody a genuine and valuable alternative to that commitment.

What we value is affected by what we take the candidates for value to be, and the values that we endorse may well change as the candidates that we consider change. This point is made over and over again by those concerned with the possibility of adaptive preference: we should be skeptical of seemingly marginalizing preferences in those who have only ever been exposed to marginalizing circumstances, because those preferences may no longer be endorsed when better circumstances become live options. It is mistaken, then, to index autonomy to exposure to the options that people themselves in fact prefer, rather than to the options that they might prefer upon exposure. If I find myself in circumstances where there is strong social pressure for women to remain in the home and care for children, and develop the corresponding preference to do so, my preference is unlikely to become much more autonomous for being considered in different circumstances where women working outside of the home are routinely attacked and beaten. In the latter case, the worse circumstances only provide more incentive for me to maintain my current preference. But my preference would become more autonomous if I reconsidered it after a brief period of interesting paid employment – and this will be the case whether or not I initially wanted to have that additional option. Because the option can reasonably be taken to be a valuable one from an external position, it provides a genuine alternative against which my original preference can be considered. So if Christman does endorse the idea that critical reflection need take place only against the backdrop of other alternatives already valued by a person, then his position must be rejected in favor of one that


301 Ibid., p. 122.

302 Of course, responses to circumstances are not mechanical, and they might have surprising outcomes. The attack of women may outrage me because I feel that such assault interferes with my choice. This, in turn, might lead me to question whether social discouragement unjustly plays the same role in my own circumstances. Generally, however, we cannot rely on consideration of worse circumstances to lead to rejection of one's own circumstances.
substitutes some more objective criterion for the kinds of alternatives that must be available.

But mere exposure to an alternative may not always be enough to inspire genuine reflection. Christman recognizes this fact and endorses it: for him, many of our commitments are ones that it is appropriate not to genuinely consider contravening. He writes that “the fact that a characteristic is experienced as so central to one’s identity and value structure that changing it is virtually unthinkable amounts to the fact that for the purposes of practical reason and pursuit of a life plan, such characteristics (and hence the effect they have on shaping that life plan) are unalterable.” 303 And our account of autonomy, he thinks, should take account of this fact: we should treat as autonomous those convictions that people cannot imagine changing, without requiring them to survive genuine consideration of alternatives.

What, then, makes critical reflection adequate, if it does not require seeing oneself as having genuine alternatives? In a slightly earlier paper, Christman suggests the following test: “a person reflects adequately if she is able to realistically imagine choosing otherwise were she in a position to value sincerely that alternative position.” 304 But such a requirement will often require far too little of people. A woman who flatly disapproves of women working outside of the home could easily grant that female employees sincerely value the work that they do, while still thinking that their views are mistaken and appalling. And if she does so, then she could likely also imagine that she herself would similarly choose to find employment if she had the same (as it happens mistaken and appalling) sincere values. But this says nothing about what she herself would actually consider or see as an alternative.

Our identities certainly deserve some protection from scrutiny. No one could live a successful life if they were to constantly consider live alternatives to their deepest commitments. But not all facets of identity can be excused from scrutiny. The paradigm case of an adaptive preference involves a value that has been endorsed to the degree that a person cannot sincerely imagine living her life in a different way, and if I want to ground my account of adaptive preferences in the concept of autonomy, then my account of autonomy cannot treat every preference that a person cannot imagine changing as immune to reflection.

In the end, I take two ideas directly from Christman’s account: that the way in which a preference was formed should count towards its autonomy, and that the

history of a preference should not alone be enough to automatically rule it non-autonomous. However, I also take from Christman’s account several points that rise from my disagreement with him: that an account of autonomy should require actual rather than hypothetical reflection, that it should require that reflection take place against the background of a set of alternatives whose value is not determined by the person’s actual preferences, and that it requires persons to see alternatives as genuine in some manner more robust than the one suggested by Christman.

An indirect substantive account of autonomy

The account of autonomy that I propose, like both weak substantive accounts and historical accounts, charts a middle way between substantive and content-neutral accounts of autonomy. Following content-neutral theorists, it takes seriously the idea that a plausible account of autonomy should not be incompatible with commonly-held conceptions of the good. And following substantive theorists, it takes seriously the idea that accounts of autonomy should provide the tools to guard against the internalization of marginalizing and oppressive norms. But unlike both strong and weak substantive accounts, and along with content-neutral accounts, I deny that substantive requirements should be made directly of people or their preferences. Yet at the same time, unlike historical content-neutral accounts, I deny that the conception of value utilized by an account of autonomy should rely on persons’ manifest preferences about value.

At the heart of my proposal is this idea: rather than either using substantive content to make direct requirements of persons or getting rid of substantive content altogether, my account of autonomy will use substantive content indirectly, to set requirements on the conditions under which a person’s preferences were formed. In this way, it will simultaneously avoid ruling preferences out on the basis of their content and guard against the possibility that a preference that develops as the result of constricting circumstances could be counted as autonomous.

As I argued in the last chapter, autonomy should be seen as a characteristic of preferences rather than persons. On my indirect substantive account, for a person X and a preference P, X will count as autonomous if and only if the following conditions obtain:

a) X has the general capacity for critical reflection;
b) X is continent in relation to P (that is, X can form motivating intentions to act and can also refrain from acting on the basis of P);

c) X actually engages in critical reflection on P and endorses it;

d) X’s reflection includes reflection on P in light of her other preferences;

e) X undertakes this reflection in light of recognized alternatives to P;

f) At least some of these alternatives are live; and

g) At least some of these live alternative options are valuable.

Conditions (a) and (b) are uncontroversial. They rule out the possibility of the autonomy of preferences arising from addiction or uncontrollable anger, and explain why autonomy is absent in cases, like those of small children or the severely mentally handicapped, in which persons cannot achieve any critical distance from their preferences at all. Condition (d) merely ensures that a person’s preferences are coherent. The interesting work in the account is done by the remaining conditions, which separate it both from substantive and content-neutral accounts of autonomy. These conditions, however, require further clarification: What counts as an option? Why need autonomous preferences be developed in light of options? What makes an option valuable, and what makes an option live? I will take each of these issues in turn.

**Options: what are they?**

Options are notoriously difficult to differentiate between. The world is filled with so many variables that even in the most limited circumstances our options might seem to be endless; yet, conversely, among this endless set, it can be difficult to tell where one option ends and the next begins. For instance, if I am choosing between entering medical school on the one hand and plucking one hair from my head and then entering medical school on the other, am I really facing two separate options at all? The question of how to differentiate options is a difficult one, but for my purposes, two guidelines should suffice.

First, to count as genuine options, two possibilities must be qualitatively different. Given this guiding principle, it should be clear that the two “options” in the medical school example above do not count as separate options. Something like the possibility of attending dental school instead, on the other hand, would count as a qualitatively different, and so distinct, option. Of course, many sets of options will be more qualitatively different than the original example and less qualitatively different than the modified example, and in these cases it will often be practically difficult to tell
when two options are involved rather than one or three. The concept of qualitative difference, however, is a useful conceptual guideline and will also serve as a useful practical guideline in many cases.

Second, to count as genuine options, two possibilities should be relevant to the situation at hand. The person going to medical school will face some alternatives that will not count as genuine options to medical school, such as becoming a vegetarian or wearing green pants. Here, the problem is not that the differences between the options are so small as to be trivial – rather, it is that the options, while different, have nothing to do with one another. In most circumstances, the decision of whether or not to be a vegetarian will be considered completely separately from the decision of whether or not to go to medical school (and what to do instead). There will, at least generally, be no sensible way to say that the possibilities can be weighed up against one another. Of course, once again, it will sometimes be difficult to tell whether two options are in the same relevant set. For instance, whether to go to medical school and whether to buy a car are not in the same relevant set if what is at stake is what I should do with my life. They may be in the same set, however, if what is at stake is how I can best spend the limited money in my bank account. Despite this difficulty, however, the idea of relevance will still serve as a useful guideline.

Options: why require development in light of them?
Like historical accounts of autonomy, my account makes the way in which a preference was formed relevant to its autonomy, but does so without holding the autonomy of a preference hostage to the way in which it was formed. To do so, it requires that preferences be developed in light of consideration of alternatives, so that persons may actively and thoughtfully choose their preferences rather than simply drifting into them.

Ideally, a preference will be critically reflected upon in light of alternatives while it is in the initial process of forming. It is ideal that critical reflection happen at this point because, as I argued in my discussion of children in Chapter Three, a history of living in accordance with a commitment can give us reasons to endorse that commitment that are different from those that we might have had at the outset. A preference that it would clearly initially be bad for me to have may later be good for me if I have lived my life in accordance with it for a significant period of time. The fact that it could be good for me later, however, gives me no reason to develop it now. This
is why critical reflection in light of alternatives should ideally happen when a preference is first in the process of development.

Nevertheless, an account of autonomy should not stipulate that a preference not developed in light of critical reflection on alternatives can therefore never be autonomous. Instead, it must be possible for preferences developed in non-autonomous ways to become autonomous by undergoing the right kind of process of critical reflection in light of alternatives. In part, this is to refrain from unjustifiably limiting the set of preferences according to which a person can autonomously govern her life— if a preference that was initially non-autonomous can never become autonomous, then many of the preferences that we developed as children can never be autonomous, no matter how thoughtfully we carry them into adulthood. But more importantly, this second way for a preference to achieve autonomy accords with a realistic understanding of the way in which preferences develop. Preferences do not have a pattern of development similar to the physical development of individual human beings. They do not experience a concentrated period of development and then become fixed more or less in the form that they will hold for the majority of their lifespans. Rather, their development is akin to the evolution of a species: preferences undergo concentrated periods of development when outside forces act on them and cause them to adapt.

This idea of outside forces is useful for determining how often autonomy should require critical reflection. While an account of autonomy that required too much critical reflection would prevent agents committed to acting only on autonomous preferences from ever holding a preference decisively enough to act on it, an account that required too little critical reflection would sacrifice the connection between autonomy and self-rule. On my account, when reflection is required will depend on the informed stability of a preference, which in turn will depend upon how much consideration of alternatives went into the preference's earlier development and upon the significance of new information to the preference in question. If a preference has previously been carefully considered against many alternatives and endorsed upon a great preponderance of reasons, then only the most extreme changes in circumstances will require critical reflection in order for the preference to remain autonomous. A preference unreflected-upon, or weakly endorsed, on the other hand, will require reflection when even fairly minor changes in circumstances occur.
Options: what makes them live?

Contra Christman, my account takes live options to be those options that persons could actually see themselves pursuing, given the values and ambitions that they in fact have.\textsuperscript{305} If a person must consider her preference in the context of this kind of live valuable option, then there is no danger that the preference that she settles upon in the end will simply be the result of an internalization of the idea that other options are unavailable or inappropriate – after all, at least one of her other options was one that she could actually see herself taking up. My requirement of the liveness of options, then, brings with it the benefits and leaves behind the drawbacks of the requirements of weak substantive accounts of autonomy. Like weak substantive accounts, my liveness requirement ensures that persons who are autonomous do not take themselves to be shackled by either norms or conventions to the lives that they currently lead. But by focusing on the way in which persons see their options, rather than the way in which they see themselves, my requirement extends the possibility of autonomy to those who may not value or experience the robust self-trust that weak substantive accounts require.

An option, however, need not necessarily be accessible in order to be live. By “accessible” options, I mean those that a person could successfully take up, given the set of external resources that she in fact has. Often, the liveness of an option and the accessibility of an option will come together. The fact that I know that I can successfully achieve a certain outcome will make it more likely to be live for me than an outcome that I have grave doubts about my chances of achieving. But the two categories can also clearly come apart. For the Jew who keeps kosher in New York City, the option of eating bacon is accessible but not live; for the child who grows up in a family that cannot afford higher education, the idea of attending college may be live despite her recognition that it is inaccessible. Since my account of autonomy requires liveness rather than accessibility, this student could have an autonomous preference to attend college despite the fact that the object of her preference may never be accessible – that is, that the student may well be unable to ever attend college.

Options: what makes them valuable?

I have said that some substantive account of what is valuable should be used to determine whether or not the options available for consideration are sufficient. In

\textsuperscript{305} I draw this reading of live options more or less directly from William James (1979) “The Will to Believe” in \textit{The Will to Believe and Other Essays in Popular Philosophy}, Cambridge, MA: Harvard University Press.
principle, my account of autonomy is compatible with any account of substantive value: any account could be plugged in, and the autonomy of ensuing preferences could be determined on its basis. By applying even quite a restrictive view of value indirectly, my account yields results more amenable to pluralism than those substantive accounts that would apply it directly. For me, it is only the absence of valuable live alternative options (however value is conceived) that makes a preference no longer a candidate for autonomy — there is no requirement at all on the content of the preference ultimately endorsed.

Given the discussion of perfectionism in Chapter One, however, the reader should be highly skeptical that such a solution can fully solve the problem of secondary recognition respect. Any theorist who offers a perfectionist account of value will be committed to saying that those who disagree with the items on the list have simply failed to recognize what is valuable in the world, even if they have done so autonomously. For instance, if this account attached perfectionist value to making one's own reproductive choices, then it could allow that a woman could *autonomously* prefer to allow her husband to make these choices, but it could not allow that, in doing so, she accurately captured a morally good way of life. And as I argued in Chapter One, taking on such a position at the state level (or at some level of similar influence) makes it difficult for individuals who genuinely disagree with the content of the list to retain their self-respect.

There are two alternatives to this kind of perfectionist position. The first stipulates a list of perfectionist values, but remains agnostic about the value of those items not on the list. My account of autonomy, when using such a list, would then seem to be able to say that the preference to make one's own reproductive choices was *certainly* valuable, but that the preference to allow someone else to make them might *also* be valuable. In this case, a perfectionist theorist using my account of autonomy need not make any statement about the moral value of a preference that diverged from the list. This solution is indeed friendly to pluralism about value — but its friendliness comes at the cost of making autonomy indeterminate. On this account, *any* alternative may be valuable, so every person who had any live options at all *might* be autonomous. And given that the motivation behind moving away from a content-neutral account of autonomy was to avoid problematic socialization where only disvaluable opportunities are available, this solution is undesirable.

My preferred alternative indexes substantive value to the idealized desires of individual persons. In Chapter Three, I defined subjective values as those that persons
themselves took to be valuable, in the real world, where they reasoned imperfectly and had incomplete information. Objective values were those that persons themselves would take to be valuable in an ideal world where they reasoned perfectly and had full information. My account of autonomy will define substantively valuable content in a similar way: the availability of those things that a person herself would take to be valuable given full knowledge of the relevant facts and perfect reasoning would be enough to count a preference as autonomous. In this way, it utilizes a definite account of value without either relying on a person’s subjective preferences or dictating a list that must be the same for everyone. One might object that such an account is not in fact substantive — after all, it offers no way of reporting on what counts as valuable in either the general or particular case. But while this strategy does not allow me to report on which values should count, it nevertheless, at least at the conceptual level, gives me a way of ensuring that problematic socialization is ruled out. In other words, it allows me to avoid the objection leveled against content-neutral accounts of autonomy. In any case, I will leave this difficult conceptual question for now. In the next chapter I will turn from questions about how to conceptualize adaptive preferences to questions of how to identify them in the real world, and I will have more to say there about how to identify valuable alternatives in practice.

Some objections

Both substantive and content-neutral accounts of autonomy have weaknesses that my account corrects. Content-neutral accounts do not adequately guard against the internalization of the norms that marginalize the most vulnerable groups within communities. By requiring that preferences be developed in conscious light of live valuable alternatives in order to count as autonomous, my account ensures that endorsement of seemingly marginalizing norms cannot be a matter of simple default. Instead, it must be a reasoned decision that actively considers and rejects obviously valuable alternatives. But my account also maintains the motivation behind content-neutral accounts: to ensure that members of marginalized groups are treated respectfully, as competent moral reasoners capable of directing the course of their own lives. It succeeds in showing persons secondary recognition respect by siding with

Note that I am here concerned with value rather than well-being. While there is something very strange about saying that a person’s well-being is improved by the presence of a value that she does not recognize, there is nothing strange about saying that a person may simply fail to recognize what is valuable.
content-neutral accounts in placing no restrictions on the content of the preferences that persons can autonomously hold.

Substantive accounts of autonomy, on the other hand, are overly prescriptive in their use of content, since they allow it to play a ruling-out function that precludes preferences associated with plausible conceptions of the good from counting as autonomous. Yet they do so in pursuit of the worthy goal of ensuring that preferences that are the product of social marginalization are not legitimated by being counted as autonomous. My indirect substantive account preserves this worthy goal without falling prey to the associated problem. It does so by placing *indirect* substantive requirements on the set of options that a person must consider, rather than placing those requirements directly on her (as weak substantive accounts do) or on her preferences (as strong substantive accounts do). This indirect requirement makes it the case that marginalizing circumstances are ruled out while preferences associated with any conception of the good can, at least in principle, be ruled in.

But the “at least in principle” is important. In practice, it may be difficult for the methods of reflection recommended by some conceptions of the good to harmonize with the requirements of autonomy that I give. For instance, some people might plausibly understand their religious faith to require unquestioning commitment. But on my account, an unquestioning commitment can only be autonomous if it was previously critically reflected upon in light of some meaningful alternative.

This is indeed a consequence of my argument, if perhaps an unfortunate one. I am looking for a plausible account of autonomy that can succeed in grounding an attractive account of adaptive preferences, and as I noted in Chapter Two, it is precisely these kinds of deferential preferences (whether to the authority of religion, husband, or parent) that frequently motivate the project of developing an account of adaptive preferences. To claim that these kinds of cases must be exempted from an account of adaptive preferences would simply be to beg the question against most accounts. But there is also a positive reason for including such preferences, given by the discussion of children offered in Chapter Three. While deferential lives might be good for some persons, they bring with them some important drawbacks. Given the drawbacks, the life may well only become a good one for some person in part because she has lived her life in this sort of deferential manner for a period of time and developed
an identity-conferring commitment in relationship to it.\textsuperscript{307} This does not make it the case that a child who has not yet developed this commitment has an interest in developing it – and indeed, the alternatives to such a way of life offered as superior by outsiders at least constitute reasons that the child herself should be able to consider before making such a commitment. So while my account does in practice make it difficult to hold some conceptions of the good autonomously, this difficulty is made necessary by the requirement of taking the needs of children as seriously as the needs of adults. And while my account does indeed rule out a full life of unquestioning devotion for the autonomous person, it leaves her the possibility of the zealotry of a convert.

My account might also seem to yield at least two other counter-intuitive outcomes. In the first case, because of its requirement of actual reflection, my account will count many of our everyday preferences as non-autonomous. Any preference that I hold but have never given much thought to, or that I have not considered since major relevant changes in my life have occurred, will be non-autonomous. I grant that many of us will be unused to seeing so many of our preferences as lacking autonomy. But why should this be? Many of the things that we do in a day are akratic, thoughtless, or a product of unendorsed habit. They are less a product of what we will for ourselves than of what we fall into by accident. If autonomy were indeed required to protect these elements of our everyday lives from paternalistic intervention, then we would have a reason to count these thoughtless preferences as autonomous. But having argued against the necessity of this requirement in the last chapter, I can see no reason to extend the concept of autonomy to these elements of our everyday lives.

In the second case, my account counts as non-autonomous some preferences that should seem to be strong candidates for autonomy on any account. For instance, consider the person who has a thoughtful preference not to be a rapist. There seems to be every reason to think that this preference could be autonomous – yet, strictly, on my account, it cannot be. This is because my account requires a preference to be considered in the presence of live valuable options for it to be counted as autonomous. Unfortunately, however, the only alternative to being a non-rapist is being a rapist. If we assume that being a rapist is not valuable, then I cannot consider being a non-rapist in the presence of any other valuable alternative. So in this case, and in other

\textsuperscript{307} Of course, deferential lives might be good for persons for other reasons as well. For instance, a person may benefit from a deferential life if his capacity for critical thought is so low that he can meaningfully direct the course of his life on his own.
cases where there seems to only be one live valuable option, my account seems to make it impossible to autonomously prefer that option.

Strictly, this is correct, but it is a problem with the letter of my account rather than its spirit. The requirements of my account are meant to label as problematic those instances in which persons endorse preferences only because they have failed for whatever reason to consider better alternatives. If it is truly the case that only one valuable option is possible, then my account has no reason to rule it out. We must be careful, however, about how we understand possibility in this case. If it is nomologically, conceptually, or metaphysically impossible to have another valuable option, then having one should be enough for autonomy. Social or legal impossibility, however, should not be enough to grant autonomy, since it is the need to combat precisely these kinds of limits that motivates an account of autonomy like mine. In any case, however, this kind of case will be rare – nearly all options will have at least one valuable alternative. For instance, when considering a preference against being a thief, a person can consider a blanket disapproval of theft against a disapproval of all theft excluding the “Robin Hood” variety. In such cases, my account gives us the right answer.

Autonomy and adaptive preferences

My primary project in this thesis is to provide an account of adaptive preferences. I have argued that such an account should be grounded in the concept of autonomy, and I have sketched a promising account of the latter here. But not all non-autonomous preferences will be adaptive. Identifying adaptive preferences requires making a further distinction. Although the accounts of adaptive preferences in the literature disagree on much, they all agree that the adaptiveness of a preference relates at least in part to its development in limited circumstances. Some elements of the account of autonomy that I have offered are clearly irrelevant to adaptiveness. Preferences that arise from laziness or addiction, and preferences of persons who wholly lack the capacity for critical reflection all lack autonomy, but none involve adaptation to circumstances. To offer a plausible account of adaptive preferences my account of autonomy will need to be limited along at least this dimension: only those autonomy deficits that are the result of a lack of options will count as adaptive preferences.

Consider two cases. In Case One, a student in the US is finishing her undergraduate degree. She gives little thought to her future, and her parents suggest that she get an MA in education and become a teacher. She does not find the option
repellent, and giving little thought to any other possibilities, she adopts the preference to become a teacher and applies for the MA program. Case Two involves a black high school student in apartheid South Africa. Under apartheid, becoming a teacher was one of the very few options available to black students who wanted to receive higher education. Accordingly, the black student in our example develops the preference to become a teacher, and applies to the relevant program. Both students’ preferences lack autonomy, but they are clearly different. The student in Case One has other options, but fails to consider them. The student in Case Two has no other options for pursuing the kind of life that she wants to live. Only the second case is an instance of adaptive preference.

Let us return to the conditions for autonomy in order to see which conditions might be failed because of lack of options. First, a preference can fail the test simply because meeting conditions (f) or (g) is impossible. That is, a person might simply have no live alternatives options, or none of her live alternatives may be valuable. In this case, she may reflect on her preferences, but still fail to render them autonomous. Second, a person might fail to reflect on her preferences in the first place because of her lack of live options. In this case, her preference fails condition (c), but does so because her options are limited. This is what happens when a person puts out of her mind possibilities that she would be likely to see as valuable because considering them without the possibility of achieving them would simply be too painful.

An adaptive preference, then, is a non-autonomous preference that fails to be autonomous either because it fails conditions (f) or (g), or because it fails condition (c) in a way that relates to (f) or (g). This definition, clearly, relies heavily on the concept of autonomy. What makes adaptive preferences suspect is what makes all non-autonomous preferences suspect: the lack of (or at best, compromised) self-rule that they exhibit. But what makes them problematic enough to receive attention as a separate category is the social failure to provide the kinds of opportunities that would make robust self-rule possible. Accordingly, two of the factors that lead to the non-autonomy of preferences deserve further discussion in light of the political project associated with these failures. These factors are the value and the liveness (as opposed to the availability) of alternatives. While a clear account of each of these factors is necessary for making sense of the general concept of robust self-rule, mischaracterizing either of these factors raises specific problems for the political project associated with adaptive preferences. I take each in turn.
Dale Dorsey writes about "status-quo bias" rather than using the terminology of adaptive preference, but his concerns are similar. Like adaptive preference theorists, he recognizes that persons' preferences can become skewed towards the options that they have available to them, even when those options are wanting. Accordingly, he worries that preferences that display a status-quo bias can be suspect sources of information about a person's well-being. Like me, Dorsey holds that addressing this worry requires assigning value in a way that is "abstracted from the agent's actual conception of the good ... [instead, value] must be identified with what a person would prefer under certain counterfactual conditions." But Dorsey's criteria for a preference to count as welfare relevant are extremely demanding; he requires that it be judged against a complete preference ordering that involves judgments about all possible lives given the experience of living them. Dorsey holds that such an ordering solves the problem of the status-quo bias:

Because ... welfare relevant preferences are drawn from a complete ordering of all possible lives, and because that ordering is generated by a conception of the good that is completely tested against judgments about the value of lives made on the basis of experiencing those lives, there is no room for an agent's good to be more heavily weighted towards the one she actually lives.

Now, to be sure, Dorsey is not explicitly concerned with adaptive preferences in this article. Instead, he is concerned with determining which preferences, in theory, display a bias for the status quo. Many of the examples that he uses to motivate his project, however, are drawn from the adaptive preference literature or might as well be. This suggests that Dorsey's solution to the status-quo bias is also intended as a solution to the problem of adaptive preferences. But an account of adaptive preferences that simply identifies which preferences are skewed by limited circumstances will be poorly suited to aid the political project. To contribute to such a project, an account of adaptive preferences must be able to identify, even at the level of theory, which kinds of skewed preferences warrant a social response. And it cannot be the case that all preferences that are suboptimal on a complete ordering like Dorsey's require such a response. Taking such a position would render virtually all of our preferences problematic, and would accordingly leave the political project too diluted to be useful.

309 Ibid., p. 548, italics in the original.
310 Ibid., p. 550.
Dorsey is right to hold, however, that we cannot rely simply on the preferences that persons happen to have to determine which options count as valuable. To take advantage of this observation, however, while still serving the political project, we can once again apply value requirements indirectly rather than directly. When discussing autonomy more broadly, we were able to show persons secondary recognition respect by requiring that they had some valuable options from which to choose, rather than requiring that the option that they eventually chose counted as valuable. In doing so, we were able to cede authority to persons' own processes of moral reasoning, while still ensuring that they were provided adequate environments in which to undertake that moral reasoning. In the specific case of adaptive preferences, applying value requirements to available options rather than to ultimately chosen preferences is once again useful. Here, it holds that social response is appropriate in instances in which persons lack valuable options to be preferred, but offers no support for social intervention in the case of preferences that merely happen to be less than ideal. This is crucial, since many of our preferences are less than ideal in ways that have been adapted to circumstance. And because my account requires only the recognition and consideration of some live options of roughly equal value, it does not err by counting too many of our everyday preferences as adaptive.

But if we can err by identifying too many preferences as problematic in light of the political project, we can also err by failing to identify enough. To see how this might be done, let us turn to the issue of liveness and availability of options. Donald Bruckner defines an adaptive preference as "a preference that is regimented in response to an agent's set of feasible options," and his aim is to defend the claim that adaptive preferences can be fully rational and worthy of pursuit.

I do not deny this. Adaptive preferences are generally understood in a negative light, often as preferences that persons ought not have. On my account, however, adaptive preferences need not be preferences that are bad for a person. While they do warrant heightened scrutiny, they do so on the grounds of the way in which they were formed, rather than on the grounds of their content. Accordingly, the content of an adaptive preference may be either similar or dissimilar to the content of the preference that the person would now autonomously endorse if she critically reflected upon her

---


312 As we will see in the next chapter, however, content may be sometimes be taken to be an indication that preferences were formed in problematic ways.
current preference in light of live valuable options.\textsuperscript{313} Think again of the black South African teacher. When apartheid ends, imagine that an ANC member approaches her and offers to give her either a public service job or the capital to start her own business. When the teacher reflects on these options, she might prefer to take one of these alternatives – but she might also prefer to remain a teacher. Being a teacher need not be bad for her simply because her preference was formed under conditions that inhibited her choice of alternatives. Such a preference need not always warrant the suspicion that comes with adaptiveness. On this point, Bruckner and I agree.

Where Bruckner and I differ is on the earlier point: just what it means for a preference to be adaptive in the first place. Bruckner defines a preference as adaptive only when it is “conditioned by what is and is not in an agent’s feasible set”\textsuperscript{314}, and for him, feasible options seem to be much more closely related to accessible options than to live options. He holds, for instance, that a young man who would receive significant ridicule and lack of support from friends and family for taking up a blue-collar trucking job nevertheless has that option in his feasible set. Accordingly, when he abandons the preference for trucking that he would otherwise have, Bruckner is clear that he does not do so adaptively.\textsuperscript{315} On my account, such ridicule and lack of support could easily prevent the option from being \textit{live} for this young man, and would therefore contribute to the adoption of an alternative preference counting as adaptive.

This difference between our accounts is clearly important for the political project. Bruckner may be right that it is rational and good to act on adaptive preferences when there are genuinely no better options that exist. There would be little value in encouraging persons to pine after the things that they truly could not achieve. But adaptive preferences are most worrisome from the perspective of the political project when they cause people to fail to develop preferences for the valuable and distant possibilities that might otherwise be achieved. Discussions of adaptive preference and patriarchal tradition go hand in hand precisely because the social costs of flouting social norms are hypothesized to keep women from seeing social change as a real possibility even when it could in fact be achieved through the right kind of social action. Such preferences exist because options which are available (albeit not immediately or obviously) fail to be live, and any account of adaptive preferences that takes the political project seriously must count them as adaptive.

\textsuperscript{313} In the diagram below, I label these preferences, respectively, “autonomous preference similar” (APS) and “autonomous preference dissimilar” (APD).


\textsuperscript{315} Ibid., p. 312-313.
Bruckner does recognize that there are such moral and political reasons to care about adaptive preferences, and before he ends his paper he briefly notes that he agrees with Martha Nussbaum that “in order for an agent’s adaptively-formed preference to count [in the social choice function], morally and politically speaking, the agent’s feasible set must meet any applicable moral conditions.”316 He agrees with Nussbaum that there are such conditions, but places further consideration of what those conditions are beyond the scope of his project. For any account of adaptive preferences that takes the political project seriously, however, such conditions cannot be put to the side — instead, they must be at the very heart of the account. Furthermore, those conditions cannot focus only on the options that are actually available to an agent; instead, they must identify cause for concern in cases where persons fail to see other valuable alternatives as live and for this reason do not undertake the work that they could to make them more readily available. This should certainly serve as a basis for public policy, but it should also be a reason to encourage individuals to see new options as live, even when they are difficult to achieve.317 By exempting persons’ preferences from suspicion only after they have been considered in light of genuinely live valuable alternatives, my account does precisely this.

I end with a caveat involving a stipulation about the way in which I will use the term “adaptive preference” in the rest of the thesis. As I noted in the introduction, some theorists prefer to use the term “adaptive” to refer to any preferences that are formed as a result of their circumstance. The political project, however, will once again become too diluted if it is concerned with all preferences that have been limited by circumstance. If I grow up in a part of Asia where rice is the primary dietary staple and wheat is rare, I may come to prefer rice-based meals to wheat-based meals, and on my account, such a preference is technically adaptive since no valuable alternatives to rice were available for critical reflection.318 But the political project is clearly not concerned with preferences of this sort. Rather, it is concerned with preferences that cause people to live lives that they would otherwise have no desire to live.

316 Ib.d 11, p. 323.
317 This is the reason, briefly referenced in the last chapter, that Ben Coulburn’s autonomy-based account of adaptive preferences is problematic. (See Ben Colburn (2011) “Autonomy and Adaptive Preference” Utilitas, 23(1).) Because it counts as adaptive only those preferences that were formed “behind a person’s back”, it cannot explain what is wrong with instances in which persons engage in the kind of character planning that involves intentionally downgrading preferences for options that are not live but might become accessible with work.
318 Note that I could also have a non-autonomous preference for rice in a way that came about in the presence of live valuable options: imagine that my family tends to make rice-based meals, but that I routinely buy my lunch in a cafeteria that offers many non-rice-based options. Nevertheless, I do not try the other options, and continue to purchase rice at lunch.
Accordingly, in the rest of the thesis I will call adaptive only those preferences that are both formed as the result of a lack of options, and that are what I will call "core" preferences. A core preference is a preference about one of the most important areas of a person's life, and it is contrasted with preferences that relate to comparatively unimportant — that is, "trivial" or "non-core"— areas.319 Others may prefer to call trivial preferences of the rice variety adaptive as well, differentiating between "trivial" and "non-trivial" adaptive preferences. I have no principled objection to doing so; I merely stipulate this more limited use of the term for simplicity's sake. I do, however, take the political project to be concerned only with preferences that fall into the core category.

The preference space, fully differentiated, looks like this:

![Preference Space Diagram]

---

319 As in the case of value, there are different ways to give content to the set of preferences that count as core. Once again, I take the use of a single objective list meant to apply to all persons to be too hostile to value pluralism. Accordingly, I prefer to index coreness, like value, to the individual: those preferences that she would take to be core in light of perfect reasoning and exposure to all alternatives count as core; those that would not, do not. For practical questions involving identification of core preferences, see the next chapter.
A note on idealization of the account

Admittedly, this division of the preference space is an idealization. As I argued in the last chapter, autonomy is a scalar, rather than a binary attribute. In the real world, this scalarity will apply at many points: a preference can be subject to more or less critical reflection, it can be reflected on in light of many alternative options or few, it can be more or less trivial, and a lack of reflection can be more or less a function of limited options. So while the examples above were carefully chosen and constructed to fit into the chart and to illustrate what is at stake in determining whether a preference is autonomous or adaptive, many real world examples will fall somewhere in between boxes. Many preferences will receive only a small amount of critical reflection, or will be considered in light of one valuable alternative but no more. Still other preferences may hover in between coreness and triviality.

So, what is the minimum number of live valuable alternatives that should count a preference as sufficiently autonomous? Just how much critical reflection is required? How live must an option be? Attempting to give precise answers to these questions would, I think, be a mistake. Even if drawing a line were to be appropriate for a scalar concept like autonomy, and even if a line could be drawn non-arbitrarily in any particular case, where that line should be drawn and why would surely vary across cases. It is appropriate that thresholds and variable values will be different whether one is reflecting on choosing a career, a romantic partner, or a savings account. Given this, I will not try to construct a single comprehensive account that can deal with all preferences in grey areas. Life is complex and complicated, and to ask a theory of autonomy or adaptive preferences to offer clear-cut evaluations in all cases would be to inappropriately force a messy world into an artificially neat framework. Nevertheless, the presence of these grey areas will make a difference to the way that it is appropriate to respond to the preferences that might fall into them, and I will discuss this point more extensively in later chapters.

Conclusion

Admittedly, the conceptual account of adaptive preferences that I have so far given will be of little use at the identification stage that is so important to the political project. My conceptual account aims only to tell us which preferences count as adaptive in theory, where epistemic limitations do not apply. In practice, of course, when adaptive preferences are identified in the real world, it will often be difficult (if not impossible) to determine whether an option is valuable for a person, whether it is live, and whether
it concerns a core preference. For all of these reasons, the identification account of adaptive preferences that I offer in the next chapter will look significantly different to the conceptual account that I have offered in this one.

Nevertheless, the conceptual account offered here can already satisfy several more of the considerations for an account of adaptive preferences developed in the first half of the thesis. Perhaps most importantly, the account given so far can respond to the consideration that concerned content neutrality. This consideration was:

Consideration 9: Content neutrality. An account of adaptive preferences should not be committed to a perfectionist, essentialist account of value, because such an account of value limits the ability of states and other actors to show secondary recognition respect for individuals who do not value the content given.

My account of autonomy, by refusing to apply substantive or perfectionist content directly to a person or her preferences, ensures that the concept of adaptive preferences can be employed while still extending secondary recognition respect to persons with virtually all conceptions of the good. At the same time, however, it employs substantive content indirectly, to ensure that important preferences formed because persons take themselves to have no other options will be appropriately labeled as adaptive. In this way, it also satisfies the consideration that concerns the “double-checking mechanism”:

Consideration 9: Mechanism for “Double-checking.” An account of adaptive preferences should utilize some mechanism for ensuring that it does not yield dangerously parochial results;

as well as the consideration that concerns live options:

Consideration 5: Live options. In determining whether the absence of a particular behavior reflects an adaptive preference, an account of adaptive preference should take into account the extent to which a person sees that behavior as a live option for herself.

And finally, because it is grounded in an account of autonomy, my account of adaptive preferences makes it possible for a person with an adaptive preference to transform her preference into an autonomous – and therefore nonadaptive – one. It shows secondary recognition respect for persons by respecting their considered judgments, and allowing them to autonomously maintain whichever preferences they endorse upon informed critical reflection. In this way, it satisfies the consideration that concerns the transformation of preferences:
Consideration 6: Transformation of preference. In order to show secondary recognition respect to persons, an account of adaptive preferences should allow for the possibility that some preferences previously counted as adaptive may become nonadaptive without any change in their content.
Chapter Six: 
Identifying Adaptive Preferences

In this chapter, I present a method of identifying adaptive preferences that utilizes substantive lists. The lists that I propose, however, are different from the lists proposed by theorists like Martha Nussbaum and Serene Khader. Rather than relying on a single list, I propose two different types of lists that correspond to the various elements of adaptiveness that arise from my conceptual account of adaptive preferences. The Vague List has much in common with lists like Khader's and Nussbaum's, in that it relies on broad deliberative agreement to identify those vague areas of life that persons take to be most valuable. In order to show them secondary recognition respect, however, the list is only a proxy: it represents a wide consensus while recognizing that some individuals will rightly take the list to be inapplicable to their own lives. This kind of list identifies the areas that persons are likely to see as core, and leaves it up to adults to decide for themselves which options count as valuable for them in these broad areas.

I also propose the use of Specific Lists in the identification of adaptive preferences. Specific Lists allow my method of identification to better serve the political project associated with the concept of adaptive preferences, and to show appropriate concern for the needs of children. Specific lists are developed by those who have experience both within and outside of the circumstances in which children live, and aim to identify the specific practices and opportunities that children would be likely to see as valuable on an informed desire account. These lists help to identify the instances in which children risk developing adaptive preferences by failing to be exposed to these valuable alternatives, and allow the possibility of response aimed at these cases.

Two problems at the identification level

At this point, we have the following definition of adaptive preferences:

A preference is adaptive if and only if it is a core indirect substantive autonomy deficit caused by a lack of options.
But while this definition is an attractive one at the conceptual level, for the reasons offered in previous chapters, it has two *prima facie* drawbacks when considered as a basis for a broader theory of adaptive preferences.

First, it may strike the reader as a poor tool for determining whether any particular preference in the actual world is adaptive or not. The conceptual account requires an account of each individual’s informed desires to determine both whether an option is valuable and whether it relates to a core area of that individual’s life. In the here and now we may generally be able to make reliable judgments about our actual desires, and may even be able to make reasonable guesses about our objective interests, but neither a person herself nor an outside observer will be able to say with certainty whether these guesses would accurately capture her actual informed desires. Similarly, the conceptual account relies on the idea of liveness, but as we saw in the last chapter, liveness can be difficult to distinguish from accessibility. It will be especially hard for an observer to determine whether a clearly accessible option is live, but it may also be difficult for a person to make this kind of judgment about herself. As we saw in the last chapter, a person who genuinely values an option may still fail to have the self-confidence or self-worth required to see it as live.

Second, my conceptual account may seem to be overly inclusive: that is, some of the preferences that it identifies as adaptive will seem to be irrelevant to the political project associated with the concept. Imagine that a privileged young man wants to be a doctor, like his father and his father’s father before him. Is this preference adaptive? According to the conceptual account, it certainly could be. He may have developed the preference to become a doctor only because he felt, unconsciously, that his family would be horribly disappointed by his taking up any other line of work. Because of this unconscious feeling, he may have never considered other options as live ones. And the young man, along with his family, might consider the choice of a profession to be one of the most important choices that a person can make. In this case the preference would certainly be adaptive, but it is clearly not the kind of case that the concept was intended to address.

At the conceptual level, we should accept both of these consequences. It is no strong objection against a theory that it cannot straightforwardly be applied to the real world or that some of its elements cannot be clearly identified in practice. Utilitarianism, for instance, tells us that the right action is the one that maximizes utility, but recognizes that it will often be difficult if not impossible to determine in advance (or perhaps even *ex post*) which action will in fact bring about the greatest
utility. What *would* count against utilitarianism (at least as an action-guiding moral theory) would be a failure to provide a justified procedure for applying its theory in the real world. This, utilitarianism can do. Recognizing epistemic limitations, it can exhort us to maximize *expected* utility within some reasonable time frame. Since my theory of adaptive preferences is meant to be response guiding, I must do something similar: I must explain how my plausible conceptual account can with some modifications also prove to be a *useful* tool for identifying adaptive preferences in practice.

Similarly, the second consequence should be accepted at the conceptual level. The concept of adaptive preferences is valuable *because* of the political project — because it is problematic that an individual's circumstances can cause her to become an instrument of her own disadvantage by preventing her from so much as desiring a life that might be better for her. And we think that such preferences require social redress because we see people who seem to have come to prefer lives that are bad for them in deep and serious ways. The cases that motivate the development of the concept are of this serious kind: they are the woman who comes to think that her husband's beatings when she disobeys him are appropriate, the lower-caste Hindu man who sincerely thinks that he does something wrong when he looks a person of higher caste in the eye, and the widow who thinks that her state of malnourishment is acceptable. Obviously, we have reason to care more deeply about these cases of greater disadvantage than we do about cases of lesser disadvantage, but while the political project was not motivated by the case of the soon-to-be doctor, that project offers no reason to deny that the doctor's preference could count as adaptive at the conceptual level. The doctor's preference does not disadvantage him relative to most other persons, as the preferences of the beaten woman, the lower-caste man, and the malnourished woman do. To the contrary, it renders him comparatively much *better* off than most people. But relative to the kind of life that the doctor himself could counterfactually have had, he may nevertheless be worse-off. It is a strength of an account of adaptive preferences, rather than a weakness, that it can recognize that it is also problematic when circumstances force a person to prefer this kind of counterfactual disadvantage, even if the person is well-off relative to those around him. Such an account should not cause us to forget, however, that this case is much less serious than the others. Accordingly, in this chapter I will seek to develop a procedure for identifying adaptive preferences that focuses on adaptive preferences of the more serious sort.
Situating and refining lists

Recognizing these limitations and constraints, I propose an identification account of adaptive preferences that relies on the use of substantive lists. This may seem surprising, given my argument in Chapter One against the use of the substantive lists proposed by Nussbaum and Khader. If I argued forcefully against the use of such lists there, why make such a similar proposal now? And even if I were to rely on a list, why propose a new one? After all, the relevant literature in philosophy and development studies already suffers from an over-proliferation of such lists. In a recent paper, Sabina Alkire contrasts more than 30 such lists in an appendix that by no means exhausts the collection of lists on offer. But while there are indeed many lists, they serve different purposes, and as I will show, the most plausible list-based approach to the identification of adaptive preferences neither replicates the work done by other lists nor shares their flaws.

The perception that the literature suffers from an over-proliferation of lists comes in large part from the failure to recognize that the lists on offer do not merely represent competing attempts to do the same work in the best way. To the contrary, they serve many different purposes, and their contents therefore vary accordingly. Ingrid Robeyns helpfully distinguishes between lists used for measurement of quality of life, thick descriptive analysis, and normative theorizing about the goods of human life. Lists used for the identification of adaptive preferences will generally fall into the third category, but even in that category there are a variety of different purposes that a list can serve. In distinguishing these purposes, it will be helpful to place lists along two axes: the perfectionist-pragmatic axis and the comprehensive-partial axis.

Where a list falls on the comprehensive-partial axis will depend upon the areas of human life that it seeks to cover. At the extreme comprehensive end, a list would seek to include literally all things, no matter how trivial, that contribute to a good human life. Monist accounts of the good, such as hedonism, will fall on this end of the scale, as would pluralist comprehensive doctrines of the good of the kind discussed by John Rawls which “cover all recognized values and virtues within one

precisely articulated system." At the extreme partial end will be lists that make claims about the good in only one very specific area of life, such as education or sexual health. Most of the lists to be found both in the philosophical and development literature will fall somewhere in the middle. All accounts of human rights make claims that cover some important areas of human life but leave many others out. Similarly, theories of political liberalism offer lists of goods owed to persons in the political realm alone, and make no further claim about the goods that matter in other parts of life. And pluralist objective list theories generally seek only to enumerate some centrally important goods, recognizing that whether or not less central items will contribute to a person’s good will depend largely on her circumstances.

Where a list falls on the perfectionist-pragmatic axis, on the other hand, will be determined by the kind of claim to truth that is made about the good of the items that populate it. If a list falls at the extreme perfectionist end of this scale, then it claims that the items that make it up are constitutive of human good. Objective-list theories of the good fall on this end of the scale, since they claim to identify those things that contribute to people’s good whether or not people themselves recognize their contribution. Lists of virtues developed by virtue ethicists will also fall at this end of the scale. If, on the other hand, a list falls at the pragmatic end of the scale, then it leaves aside the question of whether the items that it includes are genuinely constitutive of human good. Instead, it merely claims that there are some other reasons that justify treating the items on the list as relevant to human good. Items might be included on a list because they are means of promoting the constitutive

327 See Aristotle (1980) Nicomachean Ethics, D. Ross (trans.), Oxford: Oxford University Press. For more recent work, see Rosalind Hursthouse (1999) On Virtue Ethics, Oxford: Oxford University Press. Natural law theories of human rights might also fall at this end if they hold that the goods that they protect come out of our nature as human beings, although not all such accounts will. For an account of rights based in natural law theory, see John Finnis (2011) Natural Law and Natural Rights, Oxford: Oxford University Press.
goods of human life, either for an individual herself (as in the case of John Rawls's primary goods, which are not taken to have intrinsic value themselves but which can be used to acquire things that do), or for others (as, perhaps, in the case of manners, where my telling of white lies might bring about happiness in others). Alternately, they may be included because they are taken to be good proxies for getting at constitutive goods that we lack the epistemic access to identify. Expected utility based lists would fall into this group.

Natural law theories of human rights; Khader's components of human flourishing

Political liberalism; Error theories of human rights; Nussbaum's (recent) capabilities list


dot

Perfectionist

Partial

Comprehensive

Pragmatic


Figure 3: Typology of lists of values

Where on the above diagram should a list aimed at the identification of adaptive preferences be placed?

I begin with the horizontal axis. The list that I propose will be partial, rather than comprehensive, on the grounds that comprehensive lists are both difficult to make and difficult to use. Developing the kind of lists required for perfect identification on my conceptual account of adaptive preferences would be an especially difficult task: it would require developing a separate list for each individual of all valuable

---


329 I return to the vertical axis below, but I will flag here that my list will fall on the pragmatic side of the scale.
alternatives, all live alternatives, and the comparative importance of all areas of life. But while such lists could be developed in theory and with proper epistemic access, developing them in practice would be impossible. Primarily this will be impossible because we lack the kind of epistemic access that would be required to develop them. But comprehensive lists more generally will be extremely hard to populate given the almost infinite variety of things that could plausibly give value to a life. This, in turn, will make them unwieldy to use at the identification level, since checking a person’s preferences against the list would be an almost endless task. Partial lists are comparatively easy to use for identification purposes, since their use will involve consideration of only a few items whose choice has been constrained by some set of non-arbitrary guidelines.

The trick involved in developing a partial list will be choosing an appropriate set of non-arbitrary guidelines to constrain its contents. I will use as guidelines three of the overarching aims of this thesis, presented in the introduction: provision of respect for persons in the application of the concept, usefulness for the political project that motivates the use of the concept, and suitability for the needs of children as well as adults.

**Guideline One: Showing respect through the use of non-perfectionist proxy lists**

Since primary recognition respect concerns the extent to which persons are able to live their lives in accordance with their convictions – that is, what persons are able to do – I leave discussion of primary recognition respect for the next chapter where I discuss appropriate responses to adaptive preferences. For now, I will focus on avoiding the kinds of secondary-recognition-respect-related mistakes that could be made by a list used in the identification of adaptive preferences. Although Nussbaum’s list of capabilities was not designed primarily for the identification of adaptive preferences, it will be useful to consider it here for two reasons. First, Nussbaum’s primary project of developing a list of centrally important human capabilities is heavily influenced by concerns about adaptive preference. Second and relatedly, many of the secondary-recognition-respect-related charges aimed at her account are aimed at the elements of the account that respond to concerns about adaptive preferences.

In keeping with her broader concern with pluralism, Nussbaum holds that her list of capabilities should not be a culturally specific relic of those who develop it. Instead, she holds that it is “very important that people from a wide variety of cultures
... should agree that this list is a good one, one that they would choose,"330 because this agreement will indicate both that the contents of the list are likely to be correct, and that the list will be one that can command support from a variety of people. Commanding such support is essential if the list is to serve as a basis for the kind of international political consensus for which she aims. Accordingly, recognizing that the particulars of her own list are fallible, she repeatedly insists that the list that she presents “might be wrong in detail” and that its content represents merely a “proposal for further debate and argument”331 aimed at improving the contents of the list. But there are, she holds, limits on what this sort of debate can achieve. Noting the problem of adaptive preferences, Nussbaum argues that we can only have an adequate theory of social justice “if we are willing to make claims about fundamental entitlements that are to some extent independent of the preferences that people happen to have, preferences shaped, often, by unjust background conditions.”332 Accordingly, she stipulates that “a habituated preference not to have one of the items on the list... will not count in the social choice function.”333

The important criticism is this: by granting philosophers the right to decide when persons’ preferences should be discounted in discussions about the good, Nussbaum’s method effectively grants philosophers the authority to judge an individual’s preferences as illegitimate. This criticism comes across clearly in Alison Jaggar’s work. Jaggar objects to the prerogative which Nussbaum grants to philosophers of ignoring those preferences that contradict Nussbaum’s list, arguing that by “explicitly authoriz[ing] philosophers to dismiss the ideas of those with whom they disagree [Nussbaum] manifest[s] an illiberal disregard for others.”334 In discussions on the fraught subject of well-being, Jaggar worries that Nussbaum’s method of moral reasoning privileges the voices of those who are already powerful, dismissing outright the voices of those with less power. Those whose desires are judged to be, in Jaggar’s words, “corrupt” or “mistaken,”335 do not even have an opportunity to defend the values that matter to them. Instead, the powerful are

335 Ibid., p. 318.
entitled to stipulate from the outset which values persons must necessarily be wrong
to endorse.

Jaggar is right to criticize this element of Nussbaum’s work. It does not, in the
terms that I have been using, extend to persons the secondary recognition respect that
they are owed, since it does not recognize them as competent to make moral
judgments about which values they themselves ought to endorse. This problem will be
solved in part by placing any list used for the identification of adaptive preferences at
the pragmatic, rather than the perfectionist end of the vertical axis. If the list is meant
only to be a proxy for the kinds of things that persons are likely to find valuable when
relatively well-informed, then it need make no claims about the true value of any
preference, practice, or norm. Accordingly, it can provide useful information about the
preferences that persons would be expected to have on average, without claiming that
those who disagree with the list must necessarily be corrupt or mistaken about their
own good.

Using a proxy list, however, is not a full solution. Even a non-perfectionist list
that self-consciously recognizes that its contents will in some cases be mistaken will be
problematic if that content baselessly privileges some conceptions of the good and
ways of life over others. The tendency for this kind of baseless privileging seems to be
what motivates Jaggar not to use the concept of adaptive preference, since in at least
some places she expresses openness to the possibility that people’s preferences can
adapt in ways that cause them to support unjust circumstances.336 Her worry seems to
be that making judgments about adaptive preferences from a particular standpoint and
then applying the term to others is generally a way of failing to “take[...e] seriously the
possibility that ‘our’ own preferences might be corrupt or mistaken.”337

Once again, this concern is a valid one. It is indeed presumptuous and
patronizing to claim that only the preferences of persons in the developing world or
living in so-called “traditional societies” could be adaptive. I want to emphasize that
my own conceptual account does not make this claim. It is plausible that many abused
women in developed countries who remain with their partners do so in part because
they cannot consider other genuinely live alternatives to doing so. It also seems
plausible that the norms of beauty that obtain in the West do so because pressures to
conform make it hard to consider meaningful live alternatives. Indeed, as I noted at the

336 See, for instance, Alison M. Jaggar (2005) “‘Saving Amina’: Global Justice for Women and
beginning of this chapter, even the preference to take up a lucrative and privileged career like medicine could be an adaptive one. But recognizing that adaptive preferences can occur anywhere makes it no easier to develop the kind of list that could help in identifying adaptive preferences. Indeed, if anything, it makes it harder. In recognizing that virtually all preferences could perhaps be adaptive, we widen rather than narrow the set of preferences that are candidates for adaptive-preference-related responses.

Taking Jaggar's concern seriously, then, raises a problem for the very idea of using a list in the identification of adaptive preferences. A list by its nature will include some items and exclude others, and a list cannot simply be plucked from the air. It must be made by some group of persons. If some voices are included in the making of the list while others are not, then the list runs the risk of privileging some conceptions of the good over others without sufficient procedural justification. Yet at the same time, Nussbaum is right to note the danger of including those with adaptive preferences in the discussion that determines the list: if a large number of persons with adaptive preferences are included, then the list may be skewed in such a way that it encourages the replication of the adaptive preferences in question. At this point, we seem to have encountered a problem of circularity: to be helpful in the identification of adaptive preferences, a list should not be developed by those whose preferences are adaptive — but this seems to suggest that a reliable way of identifying adaptive preferences cannot be developed until a reliable way of identifying adaptive preferences has already been developed!

Guideline Two: Usefulness for the political project and the problems with vague and specific lists

For the moment, I want to put aside the thorny issue of who should be involved in the creation of a list, and instead discuss the second guideline for constraining the list's content: that the list, while showing respect for persons, should also be a tool for the political project. To be a useful tool, it must help to determine which preferences require a response. But a list that can pick out particular preferences as adaptive must be a specific one, and this leads to a tension in identification between inclusiveness and usefulness. This tension should become clear from the consideration of the lists proposed by both Nussbaum and Khader. Both theorists want the lists that they develop to apply cross-culturally, and both recognize that people with very different ideas about the good will be more likely to agree at the level of broad principle than at
the level of specific practice. Accordingly, they both adopt lists with very generally specified contents.

But with this solution, the tension arises. On the one hand, the more generally a list is specified, the less controversial its contents will be. But on the other, more general lists will also serve as much less useful tools for the identification of adaptive preferences, since it will often be very hard to tell whether the general item on the list is instantiated in a particular instance. For instance, Nussbaum argues that by including the opportunity for sexual satisfaction under the good of bodily integrity, she precludes the permissibility of female genital cutting “even when practiced by adults without coercion.” The problem with this claim, however, is that even the good of opportunity for sexual satisfaction, specified as it is at only a moderately general level, is too imprecisely specified to justify her claim that genital cutting is impermissible. A woman who has undergone even quite extensive genital cutting may take herself to retain many opportunities for sexual satisfaction and so take the practice to be unconstrained by the items on Nussbaum’s list. To really rule the practice out, Nussbaum’s list would have to be specified more particularly, to include, for instance, the opportunity to have unmodified genitals — but this, of course, would be a highly controversial claim both for women who undergo genital cutting in “traditional” cultures and for the many men and women in both “traditional” and “modern” cultures who undergo circumcision or other genital modifications for a variety of religious, medical, and aesthetic reasons.

Khader recognizes the difficulty that a vaguely described list poses for identification, but she treats it as a strength of her proposed account. Khader writes that, when working with a vague conception of flourishing, “it will not always be obvious whether people’s preferences are consistent with their flourishing, and there will not be a single, obvious, strategy for transforming adaptive preferences into preferences more compatible with flourishing.” This is a strength because absent clear-cut judgments and obvious response strategies, persons whose preferences were counted as suspect can themselves become responsible for thoughtfully judging their own preferences and deciding what to do about them should they prove to be non-

---


conducive to flourishing. This strategy involves external intervention, but it is ultimately aimed at empowering persons to judge and improve their own lives.

It is important to point out that Khader rolls the identification of and response to adaptive preferences into one deliberative process. I, however, am concerned in this chapter with identification, and the extent to which a list and strategy like Khader's can count as a way of identifying adaptive preferences depends on what role substantive value plays in the conceptual account of adaptive preferences adopted. A list and deliberative strategy can identify adaptiveness for Khader because on her account it is a necessary condition of being adaptive that a preference fails to serve some independently determined element of human flourishing. Outsiders can use the broad and vaguely specified list to identify those preferences that seem likely to fail in this way, and the persons whose preferences are in question can then use the deliberative strategy in conjunction with the list to determine whether or not these outside suspicions are warranted. The contents of the list must be applied by persons themselves, but the list itself nevertheless determines which preferences are adaptive.

On my own conceptual account this strategy will be much less useful as a means of identification. Recall that whether a preference is adaptive on my conceptual account is completely independent of the preference's content. Instead, its adaptiveness depends on the content of the alternative options that were present while the preference was being developed. A vague pragmatic proxy list could perhaps be used here, but it would be used much more indirectly. If we supposed that persons would generally develop preferences that were good for them when good options were available, then a preference for something that contravened the list might be suspected of adaptiveness on the grounds that the list is comprised of those things that persons generally take to be good for themselves when they are reasonably well-informed. This, however, would do more to determine whether a person's preferences are statistically likely to be bad for her, than to determine whether they were adaptive. Using the list to determine the latter would require asking persons to apply its contents to the alternatives that they had available to them when their preferences were being developed. When there were fewer alternatives that corresponded to the list available to a person while her preferences were developing, then her preferences would count as more likely to be adaptive. But this will be a very tenuous use of the list, and the realities of memory and introspection are likely to make it an ineffective one. Memories fade, and persons may well misremember the options that they had in earlier moments. More problematically, people may be loath to count their own
preferences as adaptive, and may accordingly be motivated to find support for their current preferences in much the same way that people are motivated to find evidence that confirms their current beliefs. In the latter cases, persons might be inclined to recall that they had reflected on their preference and chosen it against alternatives even if this was not in fact the case.

So while it is conceivable that a list of this sort might play some role in the identification of adaptive preferences, it seems unlikely that it will be a particularly useful one. Indeed, given how indirect a role it would play, using the list in this way would seem to miss the whole point of the political project: it would treat the extent to which people's preferences were good for them as a tool for determining whether they were adaptive, rather than using the adaptiveness of a preference as a way of questioning whether that preference is in fact good for the person who holds it. Given that the political project is ultimately concerned with ensuring that people are free to develop preferences for things that are good for them, it would seem to be more useful to skip the identification step entirely and to follow Khader directly in using a vague list and deliberative strategy as a method of response to adaptive preferences. I will save my full discussion of response to adaptive preferences for the next chapter, but it is worth mentioning at this point that Khader's response strategy could be adopted wholesale by my own account, if her perfectionist list were replaced by a pragmatic proxy list. Outsiders could use a vaguely specified list to pick out those preferences that they suspect would have been unlikely to develop in the presence of better alternatives, and could present that same list to persons suspected of having adaptive preferences. The list could then give these persons an impetus to reconsider their current preferences and provide them with new possibilities to entertain instead.

Why not, then, simply forget about the identification step entirely and adopt a modified version of Khader's response method? The first reason has once again to do with response. While introduction of a vague list may help people to reconsider their preferences, the vague options on that list may not be enough to help persons to see as live those concrete options that would count as most valuable to them on an informed desire account. As I argued in Chapter Two, nearly everyone will be aware that lives


165
lived in other places are lived differently from their own, but implicit bias, stereotype threat, and community pressure can all stand in the way of seeing these very different alternatives as truly live ones. All of these factors may contribute to any newly live alternatives being conservative, tentative, and far from the options that would count as most valuable for a person with genuinely informed desires. Such a strategy, then, would provide only a weak response to adaptive preferences, since adaptive preferences must be considered against options that are both valuable and live in order to become non-adaptive.

But while such a strategy may not provide individuals with live valuable options, it may nevertheless be the best option for response interventions aimed at adults. For a preference to become non-adaptive, it must be considered against live valuable alternatives. Any specific proxy list that is made, however, will get the value of many alternatives wrong. As individuals’ lives progress and their potential futures narrow, their histories with their commitments will influence their interests in ways that make it harder and harder to tell which alternative options will actually count as valuable for them. The more experiences and commitments a person has, and the more choices she must make about her life, the harder it will become to develop from the outside a specific list of valuable options that accurately captures her interests. Using this kind of specific and fallible proxy list in response to adaptive preferences will be highly problematic, since direct efforts to get persons to consider a specific non-valuable alternative as live may actually be harmful in the following way.

Making commitments is an important part of planning a life, and part of making commitments is no longer being willing to see certain alternatives as live. This happens both when we make practical commitments (as I might if I bought a house and was no longer willing to see living in another city as a live alternative) and when we make normative commitments (as I might if I got married and accordingly refused to consider other romantic partners as alternatives). Pressing adults who have made these kinds of commitments into seeing alternative options as genuinely live might jeopardize a part of their lives that they took to be deeply valuable. John Christman offers a defense of this position, arguing that some values are too central to a person’s identity to be open to revision. He offers the example of the harm done when one spouse tells the other that she has questioned their relationship and has found that she is indeed committed to it.342 Examples of religious faith could fall into the same

category: a devout religious believer might hold that critical evaluation of her beliefs was itself a failure of faith. Some people do have these sorts of commitments, and they may be right even on an informed desire account that questioning those commitments would be all-things-considered harmful. And since outside actors have even less access to persons’ informed desires than persons themselves, they should be hesitant to use strong measures to encourage a person to see any potentially harmful alternative as live. Accordingly, discussion based on a vague list of things that people generally take to be good may be the best response that can be made to suspected adaptive preferences in adults. It can open up a dialogue that may help some persons to consider alternatives as newly live, without placing pressure on those for whom doing so would be potentially catastrophic.

But one guideline for constraining the content of a list remains: it must be sensitive to the needs of both children and adults. And this guideline counts definitively against wholesale adoption of a response strategy like Khader’s. While a deliberative response based on a vague list may be appropriate for adults, it is not appropriate for children. By the time that children become old enough to engage in meaningful deliberation about values, their identities will have already been passively formed by the options that surround them. It is here that a specific list can be useful: it can help to identify those adaptive preferences that children are in the process of forming.

Guideline Three: Doing justice to the needs of children by constructing multiple lists

Remember that on my account, a preference must have three features in order to be adaptive: it must be a core preference, and it must have been formed without reflection on alternatives that were both live and valuable. Lists will be useful in getting at the first and last of these considerations, respectively. In the remainder of the chapter, I will outline two separate kinds of lists and clarify which of them can be applied to adults and which may only be applied to children. In what follows, however, I will not

343 Importantly, this is not to say that critical scrutiny is necessarily incompatible with either a healthy marriage or religious faith. But the requirements of marriage and faith are things about which reasonable people can disagree. All that needs to be the case for Christman’s point to hold is that some people legitimately feel that their marriage, or faith, or some other centrally important part of their lives will be seriously harmed by taking it as one option to be weighed against others.

344 Recall as well the discussion in Chapter Three of Edna Ullmann-Margalit’s distinction between opting and choosing, and the distress and disorientation that can come with opting to adopt an entirely new identity and set of commitments.
try to give content to either of these sorts of lists. Like both Jaggar and Khader, I think it important that the contents of lists used in normative theorizing not be stipulated by philosophers, but be developed in more participatory and inclusive ways. Accordingly, like Khader, I will refrain from taking speculative stabs at the content that the lists I propose might include in their finished forms. Instead, I will focus my discussion on providing guidelines for the procedures by which each type of list ought to be formed.

The first kind of list that I propose is the Vague List, which aims at identifying which of a person's preferences are core. This list will have much in common with the lists proposed by Khader, Nussbaum, and Brooke Ackerly – with one important exception: unlike their perfectionist lists, my list will be pragmatic. It will not make claims about "truly human functioning" as Nussbaum's list does, nor will it make the kind of claims that Khader does about "human flourishing." It will not even, in Ackerly's slightly less perfectionist terms, try to enumerate "what a human being born into this world should be able to choose to do in order to live a life worthy of being described as human." Instead, it will be a pragmatic list of those things that most people, upon informed reflection, tend to take to be the most centrally valuable areas of their own lives. The list will identify those things that are likely to be important to people according to their own informed lights by collecting proxy data on the things that most people do in fact value most deeply after discussion. The list is a pragmatic one because it only claims to identify what people tend to value when informed – it does not further claim that these things do in fact have intrinsic value, or that all people ought to value them. Since the list is a proxy that explicitly recognizes the possibility that individuals may rightly disagree with its contents in their own

349 This sort of pragmatic list could perhaps also justify the same grand uses that Nussbaum and Ackerly propose, similarly grounding a "set of options that each person in the world should have the capacity to choose" (Ackerly (2000) *Political Theory*, p. 114) or "providing a basis for central constitutional principles that citizens have a right to demand from their governments" (Nussbaum (2000) *Women and Human Development*, p. 12). It would do so on the grounds that the things most valued by the majority ought to be accessible to the minority without holding that those values should be imposed on the minority. But since the kind of list that I propose does not appeal to any idea of human flourishing or the truly human, it can serve this purpose without making any implicit claims about the reasoning capacity or the truth of the beliefs of those who disagree with its content.
cases, using such a list involves no danger of failing to show secondary recognition respect to persons who disagree with it.

Otherwise, the Vague List closely resembles those of Khader, Nussbaum, and Ackerly. Like both Khader and Ackerly, I hold that the method for arriving at this list should be a broadly inclusive one. Like Khader, I hold that this vague list should be "the result of a cross-cultural deliberative process," and like Ackerly, I hold that this process can have many inputs, including both actual deliberation on the question at hand and "collected wisdom" implicit in documents such as international human rights conventions. It should take people's desires seriously, and use deliberation as one method for ensuring that the desires that persons put forward as candidates for inclusion are at least somewhat informed. Precisely how much deliberation should be required and which parties should be included are important practical issues, and the way in which they are decided will have important theoretical implications. In particular, given that the set of persons involved in deliberation will necessarily be but a tiny fraction of the global population, it will matter very much how much distance exists between persons' initial and post-deliberative thoughts about value. If their positions have changed significantly, questions will arise about how much weight is carried by their post-deliberative consensus, given that a group differently comprised might have reached very different conclusions. Had other individuals been included in the conversation, they would likely have raised different considerations and objections, and these could lead the discussion to a different conclusion. These are important issues, but a serious discussion of them would require significant engagement with a wide body of literature in deliberative democracy and social science that I cannot undertake here. However these issues are resolved, however, I hold with Ackerly and Khader that deliberative input should directly justify the inclusion of its outputs in the list, rather than playing merely the kind of epistemic role that Nussbaum assigns to people's informed desires. If moderately informed desires directly justify the list, then philosophers and other powerful actors will be denied the right to cherry pick the concerns and commitments that will be included in the final product. When done appropriately, so that power imbalances are minimized and participants genuinely

350 Khader (2011) *Adaptive Preferences and Women's Empowerment*, p. 20
351 Ackerly (2000) *Political Theory*, p. 113-114. It may also be useful to include the results of social science surveys that investigate the correlations between persons' actual behaviours and their reported levels of happiness.
352 Nussbaum (2000) *Women and Human Development*, p. 153. Nussbaum also holds that deliberative agreement on the content of her list provides ancillary justification because general agreement on a set of capabilities suggests that that set can serve as a stable ground for governance.
consider the viewpoints of their fellows, developing a list through an inclusive deliberative process will ensure, in Ackerly’s words, that “all of the silent voices” are heard.\footnote{Ackerly (2000) *Political Theory*, p. 110, italics added.}

Including all of these silent voices should not, however, mean risking Nussbaum’s feared possibility that the contents of a list will be skewed by the adaptive preferences of those who engage in making it. As the name suggests, and like both Nussbaum and Khader’s lists, the items included on the first list will be only vaguely specified.\footnote{Peculiarly, Ackerly calls her list a “specific” one, which seems to me to suggest the opposite of vagueness. Ackerly (2000) *Political Theory*, p. 113. I find this strange because her list closely mirrors Nussbaum’s in both content and level of generality.} The list needs to be vague because it should aim to be universally applicable. I mean this not in the strong sense that it identify a set of items that are good for *each and every* person in the world, but in the weaker sense that it aim for a general consensus that is shared *across* the world regardless of political and social context. That is, the list will be a proxy list that can be usefully applied across the world, while still recognizing that in some specific cases, its contents may not apply. It is almost certain that no precisely specified list could be agreed upon at the global level, given the great diversity of commitments and everyday practices adopted by persons across the world, but this need not be a problem for a list of core preferences. As Khader and Nussbaum rightly note, persons can agree upon broad values without agreeing upon the particular ways in which those values are instantiated. Persons can agree, for instance, that having supportive relationships is a core part of their lives without agreeing on what constitutes a supportive relationship. Any conflict raised by the inclusion of those with adaptive preferences, then, can be solved by moving to a higher level of abstraction – if persons cannot agree, for instance, that personal autonomy is a good, then they may be able to agree that it is valuable to be free from arbitrary intervention. And since the consensus aimed for is broad rather than unanimous, small and persistent dissenting groups may be left out of the consensus altogether if necessary. Leaving them out will not suggest that their views are either corrupted or mistaken, however. The only claim made is that their preferences are not a useful proxy for the parts of life that *most people* would take to be centrally valuable upon reflection.

In addition to playing the role in response discussed above, the Vague List will also have a role to play in developing the Specific List. While the Vague List helped to identify which preferences were core, Specific Lists will help to identify which
preferences are the result of a lack of valuable alternatives. Since the political project is unconcerned with trivial adaptive preferences, the Vague List will identify the areas in which Specific Lists must be developed.

In the discussion above, I noted two ways in which a specific list might prove problematic for adults: first, it is simply likely to get the value of particular options wrong in many specific cases, given the history that persons have with their preferences; second, it may prove difficult to apply, since subjects must apply the list to options that they had in the past. For children, however, both of these problems disappear, and it is here that a Specific List can do useful work.

Take first the problem of determining past alternatives: while it will be difficult for all of the reasons mentioned above to recreate the set of alternative options that adults had in the past when their preferences were forming, focusing on the adaptive preferences that children are in the process of forming will mean engaging with a current social context. Comparing a Specific List of valuable options to the options that are accessible in a particular social context will not always be straightforward, but doing so in a current context at least allows the possibility of further investigating that context to determine which of the generally accessible options correspond to the items on the specific list.

Of course, accessible options will not always be live — or indeed, vice versa. Options that are far from accessible may nevertheless be live ones, as in the case of the young Pakistani girl who is inspired by Benazir Bhutto and dreams of becoming prime minister. Options may also be readily accessible but fail to be live, as was the case with the doctor discussed at the beginning of the chapter whose family pressured him so strongly to take up medicine that he never considered anything else. Given the ways in which these two can come apart, the most promising route for determining liveness would be to engage each person in a discussion of the options that she takes to be live for herself. Of course, persons may not initially be aware of which options they themselves take to be live, either because they have never given the issue much thought or because their sense of self undermines the possibility of pursuing the options which their values would otherwise recommend pursuing. But with enough pressing, this kind of introspection on persons’ own parts would nevertheless go some significant way towards determining both those cases in which accessible options fail to be live, and those cases in which live options existed despite a lack of accessible ones.
Such a strategy for determining liveness would not, however, do much to serve the political project, because it would do nothing to distinguish the situation of the doctor (which for many would likely yield a wealth of live options) from the kinds of situations which the concept is designed to account for: namely, those in which live options are difficult to find because of pervasive social and cultural norms. Once again, a proxy is needed, and accessibility seems to be the best candidate, imperfect though it is. It must, however, be accessibility of the right kind. It cannot simply be accessibility within a local context – instead, it must be accessibility *qua* member of a person's group. Clearly, options that are accessible for some members of a social group are inaccessible to others, as we see when opportunities are offered only to men, or to members of a certain race, or a high caste. So an option will count as adaptive in terms of liveness if it is one that is not regularly taken up by members of one's group. Not just any failure at the group level should count against liveness, however: options should be suspected of adaptiveness primarily in cases in which options identified as valuable are not taken up by persons *qua* members of a group *in which membership is unchosen*. In cases where group membership is chosen, persons may be drawn to become a member of some group precisely because they are attracted to an element of that group's membership that others take to be unattractive. For instance, a man may choose to become a monk precisely because he sees a vow of poverty as self-sacrificing. Accounts of adaptive preferences should be primarily concerned with those cases in which persons accept losses by virtue of a role that they clearly did not choose, or which they chose in the context of significant and widespread social pressure to do so. While this proxy will sometimes get things wrong, it will allow my identification strategy to focus on the cases in which the same forces make it hardest for the most people to develop informed desires.

Let us turn, then, to the second and more pressing problem with applying specific lists to adults: given the histories that adults have with their commitments, the items on a proxy list of valuable alternatives may often not apply to particular persons in particular cases – indeed, they may even be harmful. Think once again of the married person who may genuinely be made worse off by seriously considering the possibility of no longer being married. Because of the way in which her future has narrowed, considering an otherwise valuable alternative might be significantly detrimental to that which she rightly takes to be a valuable part of her life.

Once again, this issue is a problem in the context of a list applied to adults but not in the context of a list applied to children. Because children's futures have not yet
been narrowed in such a way that embracing some options requires refusing to consider others, a proxy list of options widely taken to be valuable will not risk the kinds of harms to be avoided in particular adult cases. An unmarried child who is forming a preference about her marital status does not risk doing damage to her sense of self by seriously considering the possibility of remaining unmarried as an adult. To be sure, adults around that child may already be committed to the idea that remaining single would necessarily be harmful for that child, and they may accordingly think that seriously considering that option will also be harmful. But respecting the child’s right to an open future requires allowing her to consider this question for herself when she is old enough to do so. This is why it is not problematic to apply to a child a fallible proxy list of specific goods. As long as the items on the list are widely endorsed and have reasons in their favor, she should have the option to reflect on all of them before she begins to make the kinds of choices that will narrow her future and place some of those items off limits to her.

But the biggest question still remains: if applying a list is permissible, who should make it? Here, we must have different answers for the Vague and Specific lists. Jaggar raised legitimate concerns about allowing powerful parties to decide which preferences were “corrupted” or “mistaken” and ought therefore to be counted as illegitimate inputs into reasoning about the good. While these criticisms deserve to be taken seriously, I noted that they need not arise in the context of the formation of a Vague List if that list is meant to represent a broad rather than a unanimous consensus. If the list is vague and represents a broad consensus, then harmful adaptive preferences can be neutralized in the list-making project by moving to a higher level of abstraction or leaving out those preferences that belong to a particularly small minority.

Unfortunately, this mitigating circumstance does not apply to the development of a Specific List. The Specific List serves the same role for children that deliberation does for adults: it identifies the concrete instantiations of items on the Vague List that count as valuable alternatives for the persons to whom it applies. But as we saw, concrete and specific agreement cannot be reached at the global level, given the vastly different ways in which concrete values can be instantiated in different social contexts. A Specific List must be made in a specific context – but which one? On the one hand, it seems attractive to allow the adults in a child’s community to make the list that applies to her. While very young children do not yet have a history with their commitments that will appropriately narrow their future, their informed desires would still be likely
to favor their local communities in at least some ways. While we might aim for children to have open futures, every child will be born into a specific cultural and social setting. Although some children with options will want to leave, others will want to stay, and being a part of a community means adopting at least some of its values. It might therefore seem promising to leave the formation of Specific Lists up to the adult members of a child’s own community, since they will be likely to have a particular appreciation of the value of the practices and norms that obtain there.355

Yet the point of the political project is to diminish (or rectify the harmful influence of) those instances in which people’s ambitions and values are constrained by the possibilities that they see around them, and the more locally a list is made, the more likely it is that the outcomes of the list will be provincial and limited. If adaptive preferences are a function of limited alternatives, then a failure to take those alternatives to be valuable is most likely to be replicated in a list that is made by “insiders”356 in a community where those opportunities are lacking.357

To solve this dilemma, the making of particular lists should be assigned to what Ackerly calls “multi-sited critics”.358 As we saw in Chapter Two, many theorists recognize the benefit of including “outsiders” – that is, those who are not a part of a local community and accordingly are familiar with alternative ways of doing things – in local deliberation about values.359 Ackerly’s multi-sited critics are both insiders and outsiders at once: because they are (or were at one point) members of the community in question, they have an intimate appreciation of the values to be found there; but

---

355 The term “community” is of course a vague one, which can refer to anything from a country to a village to a cultural or religious subgroup within a larger group. I intentionally employ this vague term, because the extension of the community in question will and should vary depending upon the content of the Specific List in question.

356 I borrow the term “insiders” (along with the accompanying term “outsiders”) from David Crocker (1991) “Insiders and Outsiders in International Development” Ethics and International Affairs, 5(1). An insider is a person who is “counted, recognized, or accepted by himself/herself and the other group members, as belonging to the group” (p. 155). An outsider is a person who is not counted, recognized, or accepted in this way.

357 Again, this will not always be the case. Standpoint theory rightly notes that those who suffer oppression will sometimes have better epistemic access to the harms of that oppression than those who view it from the outside. See, for instance, Alison Wylie (2003) “Why Standpoint Matters” in Robert Figueroa and Sandra Harding (Eds.), Science and Other Cultures: Issues in Philosophies of Science and Technology, New York & London: Routledge. While I recognize this point, it may also be the case that the adaptive preferences caused by oppression cloud person’s epistemic access to their own situations. My proposed solution can both take advantage of the improved epistemic access that living under oppression may bring, while also guarding against the possibility that adaptive preferences will lead to the creation of a parochial list.


because they have spent some significant amount of time living outside of the community, they also have some critical distance from which to judge it.

Some theorists will likely be critical of the proposal to leave the decision in the hands of these multi-sited critics. After all, especially in poor communities, those who have had the chance to live in other social contexts will likely be the more privileged members of the community. Susan Moller Okin criticizes Nussbaum’s capabilities list for a very similar reason. Okin argues that the priorities of actual poor women are much more basic than the ones that Nussbaum includes on her list of capabilities, noting that “the more sophisticated, even fanciful items on her list” seem to have far more to do with Nussbaum’s own priorities, as a “highly educated, artistically inclined, self-consciously and voluntarily religious Western woman” than with the priorities expressed by the women in the developing world who Nussbaum claims have influenced her list. Although Nussbaum is clearly an outsider rather than a multi-sited critic in relationship to these women, the same criticism could be leveled against my own proposal to privilege the judgments of multi-sited critics.

Such criticism, however, would be mistaken. Nussbaum’s list is meant to be one that broadly enumerates the central elements of human functioning. The Specific List that I propose has a much narrower purpose: it tries to get at the alternatives that children themselves would count as most valuable if they were fully informed and capable of carefully reasoning about the matter. So here it is a virtue rather than a problem that the experiences and priorities of multi-sited critics may be different from those who have only a more local experience. In this case, leaving the list in the hands of those with a broader variety of experiences is justified, since they can better approximate the hypothetical fully informed judgment of children for whom they are acting as proxies.

Admittedly, the value judgments of multi-sited critics will not be a perfect proxy for what children themselves would take to be valuable on an informed desire account. Each adult multi-sited critic will have her own history and set of commitments, and these will influence the value judgments that she makes in ways that could not bear on the value judgments of children. Indeed, there may even be significant disagreement amongst the multi-sited critics engaged in discussion, since their histories and experiences may vary so significantly. For this reason, deliberation will have an especially important role to play in the formation of Specific Lists. Such deliberation

will allow multi-sited critics with very different experiences to reconsider those experiences and to separate what they take to be intrinsically valuable about the options they advocate from the incidental features of those options. They can then work together to determine whether a variety of concrete options displaying those more central values are available in a particular context. When Specific Lists are the object of broader agreement by a larger and more varied group of multi-sited critics, they will serve as more useful tools in response to adaptive preferences.\(^{351}\) At the level of identification, however, it is enough for a group of multi-sited critics to agree that the valuable options of children are limited in some circumstance, even if they disagree about what the absent valuable options might be. Responding to adaptive preferences requires knowing which valuable alternatives should be introduced; identifying them only requires knowing that valuable alternatives are generally absent.

The entire process of the development of Vague and Specific lists, then, will look something like this. Initially, a world-wide discussion of core values will take place. This discussion will resemble the ones advocated by Nussbaum, Ackerly, and Khader, and will involve deliberative inquiry by a diverse range of participants about which elements of their lives they take to be most centrally valuable. This process of discussion can be on-going, and the items on it can be revised as time passes if the consensus among participants recommends it. Once this Vague List has been made, each item on the list becomes a candidate for the formation of a Specific List applicable in a certain context. At this point, a collection of multi-sited critics is brought together to consider which options would be likely to count as valuable for children in that particular context, and whether a broad range of options is indeed available. For instance, multi-sited critics might be tasked with constructing a Specific List of valuable alternatives in the area of spiritual commitment, for the children in an ultra-Orthodox Jewish community. The group of multi-sited critics might include (among others) those who were raised outside of the community and joined it later in life, those who had left the community but were still involved with it in some way, or those who had left the community and no longer maintained strong ties with it. In the best case, these multi-sited critics would agree on a varied set of options that counted as valuable, and would agree on whether or not a significant range of these options seemed to be live for children in the community. In the worst case, they would be unable to agree on which options were to count as valuable, but would at least be able

\(^{351}\) I return to this point in the next chapter.

176
to agree on whether a broad range of the options that they individually counted as valuable seemed to be live for children in the community. (This might occur, for instance, if the group included those who had converted to ultra-Orthodox Judaism to join the community, those who had left it in favor of an evangelical sect of Christianity that placed heavy emphasis on conversion, and those who had rejected the community's religious beliefs in favor of atheism or agnosticism.) Where many qualitatively-different options seem likely to be genuinely live for children, there should be little concern that their preferences are being formed adaptively. Where few options seem to be live, such concerns will arise, and response to adaptive preferences will be appropriate. I discuss appropriate responses to adaptive preferences in the final chapter.

Conclusion

The identification account that I have offered is far from perfect. The lists and methods that it employs are rough ones that rely on fallible proxies to determine which preferences will be identified as adaptive. Sometimes, the method that I propose will get things wrong. This potential for error, however, is what allows the identification account to strike an appropriate balance between showing persons recognition respect and serving as a useful tool for the political project that animates the concept of adaptive preferences. As I have repeatedly argued, only a non-perfectionist account of adaptive preferences – that is, an account that does not rely on any conception of human flourishing – can show persons secondary recognition respect, and it is therefore necessary that any external instance of identification of adaptiveness should be defeasible. Yet if a theory of adaptive preferences is to enable a response to actual adaptive preferences as the political project requires it to do, then it must have some effective way of identifying adaptive preferences in concrete circumstances.

In striking this balance, my identification account satisfies two more of the considerations identified in the first half of the thesis. In the first case, it recognizes in accordance with Consideration 10 that an account of adaptive preferences must treat the preferences of children and adults differently:

Consideration 10: Children. An account of adaptive preferences must be sensitive both to the adaptive preferences that adults already have and the adaptive preferences that children might acquire. Currently held adaptive preferences should be treated differently from those that might yet form.
My account recognizes that even adaptive preferences can come to form an important part of persons' identities, and can be autonomously maintained even once they have been adaptively formed. Accordingly, it recognizes that, in the case of adults, the content of a person's preferences will generally be a bad indication of their adaptiveness. It does not, however, entirely reject the use of substantive content. Instead, it recognizes that children's identities have not yet been formed by a history with their commitments, and so uses substantive content to identify those cases in which children's preferences are being formed adaptively. And recognizing that substantive content must be chosen by some agent, it leaves the selection of this content up to the multi-sited critics who have an experience broad enough to ensure that the content that they select will not be dangerously parochial. This allows it to satisfy Condition 9:

**Consideration 9: Mechanism for “Double-checking.”** An account of adaptive preferences should utilize some mechanism for ensuring that it does not yield dangerously parochial results.
Chapter Seven: Responding to Adaptive Preferences

In this chapter, I turn to the question of response to adaptive preferences. While many theorists champion deliberative responses to adaptive preference, and rightly fear the possibility of coercive responses, I discuss a third way: incentive-based response. While deliberative responses might be more appropriate when only adults are involved, I argue that a response that combines deliberation and incentives does a better job of recognizing and serving the interests of children. Such a response extends the range of options that children can see as live, while still showing adults the primary recognition respect that they are owed.

Expanding autonomy: coercion, persuasion, incentive

On my conceptual account, adaptive preferences are autonomy deficits. Since I have rejected perfectionist accounts of adaptive preferences, responding to adaptive preferences cannot be a matter of simply aiming to replace them with a new set of preferences that corresponds to an account of flourishing-compatible behavior. Instead, responding to adaptive preferences will mean aiming to make the preferences that persons currently have more autonomous. Given the account of autonomy that I propose, this could be done in two ways: either by changing people's accessible option sets in order to increase the set of live options in light of which they might reflect, or else by encouraging persons to see options as live even when they are not accessible in order to enable reflection on those.

Development theorists have primarily focused on the second of these possibilities. Khader, Sen, Crocker, Ackerly, and the other deliberative theorists who I discussed in Chapter Two all adopt such a strategy: they recommend engaging persons in discussion both of those options available and those unavailable in their communities in order to determine which options members of that group and outside supportive state or non-state actors should seek to make actually available. As I noted in the previous chapter, this possibility is in many ways an attractive one for adults.

But this is not the only way of increasing persons' autonomy: outside actors can also work to extend the set of actual behaviors and corresponding values that are taken up by persons in a community in an effort to make these options live ones for the
community’s other members. At the level of theory, Joseph Raz (who adopts an account of autonomy similar to my own) holds that outside actors can have a role to play both in encouraging the inner capacity to consider new options, and in ensuring that concrete options are there to be considered. In addition to helping persons to develop reasoning abilities and affective commitments, and to encouraging them to exercise their power of imagination, he writes that we have duties to other persons to create “an adequate range of options for [persons] to choose from.”

As I noted in the Introduction, I am interested primarily in what state and other state-like actors can do to respond to adaptive preferences. Like deliberative theorists, I hold that they can permissibly initiate a dialogue in which persons collectively lay out and reflect on the reasons for their preferences and actions. But in an effort to encourage people to take up new forms of behavior or adopt new preferences, can they do more than appeal to and highlight the existing reasons for doing so? Can they actively work to encourage persons to adopt a broader range of lifestyles when persons are not otherwise inclined or convinced by deliberation to take them up?

If these efforts are coercive, then I think that the answer is no. Coercing adults to adopt new lifestyles would clearly violate the requirements of primary recognition respect, since it would fail to recognize the interest that they have in living their lives in accordance with their own convictions. As I noted in Chapter Four, readers and others may value autonomy for its own sake, and show persons primary recognition respect accordingly. However, since I have not given an argument for why they should, and cannot assume that they do, I will focus in this chapter on well-being related arguments for showing persons primary recognition respect. On these grounds, coercive interventions are still to be discouraged. As I argued in Chapter Three, third parties do not know how a person’s history has affected her commitments, and therefore lack the epistemic access to determine what a person’s most important interests are. The kind of strong opposition that can only be overcome by coercion may not be proof that a person is correct about her own interests, but it is nevertheless enough to prevent a third party from coercing her for her own good, since it indicates that there is a significant risk that such coercion might render her worse off without justification.

But between coercion on the one hand, and rational attempts to persuade on the other, lies a third possibility: incentive. While coercion attempts to force a person to

take a certain course of action, and persuasion uses rational means to convince her that her circumstances provide her with reasons to undertake that course of action, creating incentives "does not offer reasons but instead changes the balance of them."\(^{363}\) In other words, incentives are a *non-rational* means of persuasion. Unlike coercion, they are intended to be genuinely refusable; unlike persuasion, they seek to change circumstances in order to encourage compliance, rather than simply appealing to the reasons for compliance that already exist. But not all means of changing circumstances in order to influence decisions seem to be properly described as incentives. Imagine that my younger sister tells me that she has a strong desire to join the golf club at her school, but has decided not to do so because buying a set of clubs would exhaust her savings. If I buy the clubs for her, then I am using non-rational means to change her decision about joining the golf club, since I am changing her material circumstances rather than appealing to reasons that already obtain. Buying the clubs, however, does not seem to be an instance of providing an incentive. Instead, it is an instance of providing the means to take up a course of action that my sister already considered uniquely worth pursuing absent her financial circumstances. For the purposes of this thesis, an incentive will be understood as an additional provision which is intended to be desirable enough to encourage a course of action that was not seen as intrinsically valuable or attractive enough to be the preferred course of action for its own sake if circumstances allowed. If my sister had told me that she was deciding between the golf club and the scuba club and was either indifferent or leaning towards scuba, then my offer to buy the clubs (but not scuba equipment) *would* count as offering her an incentive on the definition of incentive adopted here.

There has been little discussion of incentives in the literature on adaptive preferences, but the point is occasionally raised in passing and deserves further consideration. Susan Moller Okin addresses the use of incentives directly when she holds that it unacceptably impinges on people’s freedom for development practitioners or other outsiders to “dictate their priorities to them”, but speaks positively of the use of incentives to encourage people to “make non-traditional decisions that are likely to have a lasting positive impact.”\(^{364}\) Uma Narayan does not directly use the language of incentives, but she does hold that the state should “actively foster a number of legal and social changes that improve women’s powers and opportunities to negotiate their


encounters with [cultural] practices.” She writes that this should be done in ways that utilize “persuasion rather than coercion,” but she also recognizes that it is often rational for women to engage in dominating cultural practices because of the genuine goods that can be gained from doing so. Accordingly, we may perhaps understand her to be suggesting that this active fostering should involve the use of something like incentives to encourage women to choose the values to be gained from negotiating their encounters in new and progressive ways rather than the values to be gained from more traditional dominating practices. Doing so would change the set of goods which women might acquire through their choices, rather than simply relying on persuasion alone, and would recognize that it may be the case that neither course of action will seem obviously preferable to the women forced to choose. Whether or not this is Narayan’s intention, however, the use of incentives in the transformation of practices and related preferences is intriguing and deserves further attention.

A dilemma and an incentive-driven solution

Some anti-paternalists may think that even the use of non-coercive forms of non-rational intervention like incentives is impermissible when it comes to rational adults. They may hold a position like the following: “Paternalism towards persons with the ability to appreciate and act on good reasons is always wrong, no matter which of the ‘nonrational’ means it employs.” This position is controversial, but for the purpose of the chapter I can afford to remain agnostic on the point because a non-paternalistic justification for targeting certain incentives at adults exists: the good of children.

One of the central aims of this thesis has been to include the needs of children in the debate about adaptive preferences, and it is perhaps most important to do so here at the practical level of response to adaptive preferences. As I argued in Chapter Three, the good of adults can be hard to determine from the outside because of a lack of epistemic access to the unique ways in which their histories with their commitments have shaped their interests. Whether adaptive or not, preferences that were initially bad for persons to form can, once formed, begin to genuinely contribute to a person’s good, insofar as she values who she is and would not meaningfully be the same person without that preference or commitment. In other words, a preference adaptively

366 This phrasing belongs to Scoccia (2008) “In Defense of Hard Paternalism”, p. 353. Note that he does not himself endorse this position.
formed can sometimes be autonomously maintained later in life and can, at that point, reasonably be preferred to alternatives that would have initially counted as more valuable. But as I also argued in Chapter Three, this possibility does not count in favor of children forming adaptive preferences in the first place. If we value autonomy instrumentally, as I have argued that we should, then children have an interest in forming preferences non-adaptively even given the possibility that adaptively-formed preferences may genuinely become valuable to them once developed. And determining which options count as valuable for children is much easier than doing so for adults, because children do not yet have (and indeed, should be prevented from prematurely forming) the kinds of long-standing commitments that form their identities.

But this observation in turn led to a dilemma: taking the needs of either children or adults seriously seems to require frustrating the needs of those in the other group. On the one hand, if many of the adults in a community have formed the same adaptive preference in response to their shared social circumstances, then they may have incorporated that preference into their identities and may have an interest in maintaining it over the course of their own lives. But on the other, children growing up in those homogeneous social circumstances have an interest in being exposed to alternatives, so that those alternatives may become live to them, and allow them to form non-adaptive preferences in the first place. If the interests of adults in maintaining the identities that they have are privileged, then children fail to be exposed to alternatives, since the adults around them determine to such a great extent the options that count for them as live. But if the interests of children are privileged, and adults are accordingly pressed to change the ways in which they are living, then their own interest in maintaining the identities that they have may be violated.

In this dilemma, there exists a justification for intervention which did not exist in the case of adults alone. In the parallel adult case, adults themselves are both the end for which an intervention might be undertaken and the means by which it might be done, since their own behavior is meant to be changed for their own good. The problem in the adult case is that third parties cannot be sure what constitutes the good of those adults, since a person’s good is determined at least in part by her history with her particular commitments – a history to which third parties, of course, do not have access. Since the ends of intervention – that is, adults’ own good – are epistemically inaccessible to third parties, it is unclear what sort of interventions might be means of bringing those ends about. And without a good chance of bringing about the intended ends, the means lack justification. Even a change that third parties may suspect will be
in an adult's best interests might be harmful, and should therefore not be encouraged when it meets with resistance.

When children enter the picture, the case changes. Now, the good of children becomes the end of intervention, and the behaviors of adults become the means by which that end can be met. As I argued in Chapter Three, the content of the goods that children have an interest in being exposed to is more epistemically accessible than the corresponding content in the adult case, since children have not yet developed the kinds of commitments that constrain their range of interests in idiosyncratic ways. In the last chapter, I proposed that this content be identified in practice through a process in which multi-sited critics deliberated together to determine which options children themselves would be likely to prefer upon informed reflection. Admittedly, diverse groups of multi-sited critics may sometimes be unable to reach agreement on which options ought to count as valuable, and in such cases, intervention based on such lists will lack justification for the same reason that it did in the adult case: given that the good of children was epistemically inaccessible, intervention intended to promote their good would not have a clear target at which to aim and would accordingly stand too great a chance of missing. When a diverse group of multi-sited critics does reach agreement, however, that agreement can be treated as a proxy for the good of children, and in such cases, a legitimate justification for intervention exists.

The question still remains, however, whether that justification is overridden by the interests of adults, since they remain the means for such interventions. Adults targeted by interventions would be asked to change the way they lived their lives in order to make a broader set of possibilities live for children. But adults obviously have strong interests in the course that their own lives take, and taking those interests seriously means treating adults in ways that are neither over-demanding nor arbitrary. Targeting adults to adopt new behaviors in homogeneous circumstances, however, risks doing both. If all adults act in a way that would be unproblematic for any one of them to act individually, then it would be arbitrary to target any one of them for change – and as long as all adults have a strong interest in acting in the otherwise permissible way, it would also be over-demanding to target any one individual. So, although the good of children provides a justification for intervention into the lives of adults, that justification may not render such intervention permissible, all things considered.

Here it helps that the scope of the two possible interventions is different. When the interests of adults are the ultimate target of the intervention and resources are
limited, a justifiable adaptive preference intervention would need to be aimed at reliably making a large number of those adults better off. Otherwise, the resources would be better spent elsewhere on projects that the adults concerned had themselves endorsed. But when the interests of children are the ultimate concern of the intervention, a justifiable adaptive preference intervention need only have the immediate aim of making a far smaller group of adults better off (or at least as well off). Even a small number of adults changing their behaviors or expressed values will make a large number of children in the group better off by providing them with a broader set of live options against which to consider their own preferences. While it will differ from child to child how broadly and in which cases an option need be represented in order to become live, each additional expression of that option can only help.

This requirement is far less demanding. In Chapter Three, I argued only that having a history with an initially harmful commitment can give you a genuine interest in retaining that commitment, and that third parties cannot judge when this is the case. But each of us has many interests, and these interests can conflict with one another. While it will sometimes be the case that an individual's history with her adaptively-developed commitment will make the continuation of that commitment definitively in her interest, it may also be the case that her relationship with her commitments will give her reasons that can be outweighed by reasons pointing in the other direction. Having an interest in a commitment is not the same as having an all-things-considered interest in it.

Of course, weighing up one's competing interests requires consideration of information that is only available from the first-person perspective. In any particular individual case, a third party will be unable to say with any confidence what the outcome of a person's considered deliberation about her overall interests will be, and it will therefore be impermissible to push her to make choices in a certain way. So although a young woman who wants to be a housewife may seem to a third party to be made worse-off by depending upon her husband financially, that third party cannot know that this preference would not survive her considered reflection. Accordingly, the third party should not push her to adopt the possibility that she has rejected. But when many people make the same prima facie bad choice in a context in which few other options are available – as in cases where there is most cause to suspect adaptive preferences – third parties can permissibly suppose that at least some of these preferences might change were new options to arise. So if many young women
remained in the home because it was expected of them, it may be highly unlikely that any particular woman would take up employment were it offered to her – but highly likely that at least some of them would. And if even some persons do, it becomes easier for children in the community to see those options as live themselves.

Given all of this, incentives are a promising tool for adaptive preference interventions aimed at the good of children. Unlike coercive means of non-rational persuasion, incentives can be targeted broadly at a population. Yet unlike the rational means of persuasion that can uncontroversially be used with adults for their own good, incentives alter the reasons for action that exist in a particular context, and they can accordingly lead to a change in behavior that rational means alone could not achieve. This means that the benefits of incentives and the accompanying behaviors that go with them can be taken up selectively by members of a community: those who have strong commitments to their current values (and to the way in which those values are expressed through their behavior) can avoid taking them up; those who have less strong reasons, which may be outweighed by the introduction of incentives, can choose to take them up. This avoids the problems of coercive paternalistic intervention, since those external actors orchestrating an intervention no longer run the risk of pushing particular members of a population to take up a behavior that is intended to benefit them, but that, as a result of the commitments of those particular persons, may end up harming them instead. The choice to change one’s own behavior remains, in a meaningful way, with a person herself. Accordingly, if incentives are targeted broadly at a population and persons are able to choose not to take them up, then incentive-based interventions can be compatible with showing persons primary recognition respect.

While the use of incentives for adults is not justified by the good of adults themselves, it may nonetheless benefit them in several ways. In exposing children to new options, adults will be exposed as well, and this exposure may lead them, as well, to reconsider their own preferences. If they do so, and their preferences also become more autonomous as a result, then so much the better. Additionally, incentives may be a useful tool for those persons who would prefer at a higher order to change their behavior but fail to do so for external reasons, such as fear of reprisal from a partner. For instance, if a woman wants to access information about family planning but fears

---

367 Incentives might also provide a material safety net for those who already found the behavior attractive at the higher order but did not feel that they could take the risk involved in pursuing it. Think here of a person who would like to start her own small business but does not want to take on high-interest micro-loans or risk all of her savings to do so.

186
that her husband will not want her to do this, or wants to adopt sexual practices that will reduce her risk of contracting HIV but thinks that her partner will not, a material incentive may offer her a bargaining chip.\textsuperscript{368} If the material incentive is available only for engaging in the specific behavior, then the woman may be able to emphasize her desire for the material incentive that can benefit both herself and her partner, rather than focusing on the good itself which her husband may disapprove of.\textsuperscript{369} In this way, an incentive may change the reasons for action not only of the person who could take it up, but also for those persons who might otherwise prevent her from doing so.

There are, however, also several reasons to be cautious about incentives. The behaviors that adults will be incentivized to take up will be the ones that both are under-represented in the local community and were judged by multi-sited critics to count as valuable for children. But an option may meet these criteria and nevertheless be one that we would hesitate to encourage adults to take up. For instance, multi-sited critics may judge that Muslim girls living in a liberal society suffer from failing to see purdah as a live option, since it is under-represented there. Does this mean that the state should incentivize women in that society to begin living in purdah, in order that girls might begin to consider the possibility as live? I hold that it does not. While it might be valuable for girls to have this live option, the state has a responsibility not to treat adults as simply a means of broadening children's horizons. To avoid doing so, the state should only be allowed to incentivize those behaviors that are “fruitful”: that is, those that expand the set and variety of near-by options that a person can take up.\textsuperscript{370} Even if living in purdah is itself a valuable option, adopting it closes down more

\textsuperscript{368} For discussion of incentives for attending family-planning information sessions, as well as for broader discussion of the use of performance-based family planning incentives, see Rena Eichler, Barbara Seligman, Alix Beith, & Jenna Wright (September 2010) “Performance-based Incentives: Ensuring Voluntarism in Family Planning Initiatives” Health Systems 20/20, United States Agency for International Development. For discussion of the use of incentives in the reduction of HIV, see Mead Over (2010) Using Incentives to Prevent HIV Infections, Center for Global Development Working Paper. For more on the ethics of incentives and HIV prevention, see “The Ethics of Material Incentives for HIV Prevention”, Emerging Issues in Today's HIV Response: Debate 5, World Bank and United States Agency for International Development.

\textsuperscript{369} Additionally, if the incentive can only be accessed by the woman and is on-going, this form of income may provide her with on-going bargaining power to influence more general decisions in the family. See, for instance, Cheryl Doss (2013) Intrahousehold bargaining and resource allocation in developing countries, World Bank Policy Research Working Paper 6337.

\textsuperscript{370} This account of fruitfulness is borrowed from the more technical account offered by H.E. Baber. H.E. Baber (2010) “Worlds, Capabilities, and Well-Being” Ethical Theory and Moral Practice, 13, 377-392.
qualitatively different options than it opens up, since it prevents women from taking up a wide variety of pursuits outside of the home.\footnote{371}

Another reason for caution involves efficacy. If the use of incentives is not efficacious, then resources should not be wasted on them when they could productively be used elsewhere. Incentives are not always effective at changing behavior. Some studies, for instance, find that financial incentives lead to only short-term changes in health behaviors.\footnote{372} But even if these behaviors are maintained by persons accessing the incentives directly only as long as the incentives are maintained, if used widely enough they may contribute to the creation of a social norm in which the targeted behavior becomes a live option even for others not taking direct advantage of the incentive.\footnote{373} This, of course, is the aim in the case of children. By changing the lower-order preferences of adults enough to change their behaviors, incentives might make possible the development of different higher-order preferences in the children exposed to those adults. In any case, the extent to which incentives work and which incentives work best (since not all incentives need be monetary) are important questions that require further study before incentives for adaptive preference change can be used in practice. Such study, however, would require detailed empirical research better suited to social science than philosophy.\footnote{374} Accordingly, I leave these questions to one side.

A final issue concerns the legal and social constraints that exist in particular circumstances. There are certain risks which persons may permissibly take on their own behalf which it would be impermissible to press them into taking on. Think, for instance, of underground schools and vocational training centers run for women in Afghanistan under Taliban rule.\footnote{375} Those who ran and attended these schools faced serious consequences including bodily harm and imprisonment. While the women and girls who did so were courageous to do so of their own volition, using external means to encourage a person to unnecessarily take on this extreme level of risk would be

\footnotetext[371]{Of course, taking up one particular pursuit outside of the home will also prevent her from immediately taking up alternatives to that pursuit. Most particular outside pursuits, however, are likely to bring with them skills and experience that would aid her in moving on to alternative pursuits in the future.}
\footnotetext[373]{And indeed, incentives have proved to be quite efficacious in the long run in other cases, as they have in the case of conditional cash transfers in Latin America and elsewhere. See, for instance, Laura B. Rawlings & Gloria M. Rubio (2005) World Bank Research Observer, 20.}
\footnotetext[374]{For such empirical research involving incentives, see Abhijit Banerjee and Ester Duflo (2011) Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty, New York: Public Affairs; and Dean Karlan and Jacob Appel (2011) More Than Good Intentions: How a New Economics is Helping to Solve Global Poverty, New York: Dutton Books.}
\footnotetext[375]{See, for instance, Elaheh Rostami Povey (2003) "Women in Afghanistan: Passive victims of the borga or active social participants?" Development in Practice, 2(3).}
problematic at best. External actors have a duty of care to those in vulnerable situations that would be violated by pushing them to take on serious and foreseeable costs against their own judgment. Of course, there are always some costs associated with violating a social norm, and this will be so in any case in which incentives might permissibly be used. Accordingly, further work needs to be done both on which cases count as extreme, and where and how the line may be drawn for the permissible use of incentives in such cases. While this set of questions, unlike the last, is entirely appropriate for philosophical study, I will also set them to the side. Offering them a full treatment here would be too large a project for the moment, and so I leave it as an area for further study. In this chapter I focus on the narrower question of the permissibility of the use of incentives in less extreme cases.

There remains, however, one concern about the use of incentives which is too central to the project of developing an appropriate account of adaptive preferences to be put aside. I take up this concern in the next section.

The problem of coercive incentives

In some ways, my proposal may seem to resemble Richard Thaler and Cass Sunstein’s system of “libertarian paternalism,” which claims – I think rightly – to be an unobjectionable form of paternalism. Libertarian paternalism endorses a program of interventions that persons can easily opt out of, but that nevertheless use non-rational means to significantly improve the choices that persons make. Like Thaler and Sunstein, I begin with the idea that our preferences do not predate our social and legal contexts. What we prefer is, in large part, a function of what we take our options to be, and our values are to at least some important extent a function of the norms and standards that surround us.376 My proposal also echoes the “libertarian” element of theirs: we share “the straightforward insistence that, in general, people should be free to opt out of specified arrangements if they so choose.”377 My insistence on incentives rather than coercion is based in the idea that even persons suspected of having adaptive preferences may in fact have the preferences that best serve their own interests and should therefore have the real option of acting on those preferences.

---


There are also important differences, however, between my own project and the paternalistic element of Thaler and Sunstein's. For Thaler and Sunstein's purposes, a project is paternalistic "if it tries to influence choices in a way that will make choosers better off, as judged by themselves."\textsuperscript{378} The kind of paternalism that they defend, then, only justifies interventions into the lives of a group of people if those people themselves would endorse the desired change in a world that is not very far from our own. This definition of paternalism does much of the work of making it unobjectionable. My own project is explicitly not paternalistic towards adults: while interventions are aimed at the lives of adults, those interventions are justified not by the adults' own good, but rather by the good of the children who are influenced by them. Like Thaler and Sunstein, those who choose to take up incentives on my account are meant to do so because they think that the combination of incentive and behavior (and hopefully, at least eventually, the behavior alone) would make them better off as judged by themselves. But people who choose in this way may not think that the behaviors themselves will make them better off – and people who do not choose to take up the incentive may do so explicitly because they think that the associated behaviors are unpalatable enough that they cannot be outweighed by incentives. In other words, Thaler and Sunstein advocate the use of non-coercive tactics to non-rationally encourage members of a group to take up behaviors that they themselves would endorse. I, on the other hand, advocate the use of non-coercive tactics to non-rationally encourage members of a group to take up courses of action which I allow that many, if not most of them, may rightly take to make them worse off.

Second, Thaler and Sunstein are primarily concerned with responding to fairly superficial ways in which default rules, framing effects, inertia, and other similar phenomena can affect the choices that persons make. While these effects generally occur unconsciously and may often have significant practical effects, they are rarely ones to which persons have strong commitments, and are generally swiftly reversed when persons consider their situations consciously and carefully. The adaptive preferences with which I am concerned are of a different type. They are also often formed unconsciously and tend as well to have significant practical effects – but they also by definition concern areas of life about which persons have strong commitments (even if their preferences about the content of these areas are not always strong), and

\textsuperscript{378} Thaler and Sunstein (2008) \textit{Nudge}, p. 5, italics in original.

190
in at least a significant number of cases they will not be ones that persons would change were they to carefully and consciously reflect on the situation.

Finally, Thaler and Sunstein are concerned almost exclusively with the use of libertarian paternalism in the developed world, where persons are relatively well-off and can fairly easily afford to forgo the minimal benefits of complying with the libertarian paternalist’s aims. My account of adaptive preferences is by no means concerned exclusively with the developing world, but cases in the developing world are most often discussed in the literature, and they must accordingly be born squarely in mind when discussions about adaptive preferences are undertaken. And in the developing world, where many people live in precarious and impoverished circumstances, the material benefits that incentives provide may not be so easy to forgo. When incentives are offered for behavior change in areas where persons may have strong commitments and where they may rightly take themselves to be made worse off by engaging in the behavior to which the incentives are attached, it will be problematic if those incentives are difficult to turn down. In a word, those considering using incentives as a non-coercive means of bringing about behavior change should worry that, in the wrong circumstances, incentives can in fact become coercive.

There is little agreement in the literature about whether it is strictly possible for incentives to be coercive. Some treat it as axiomatic that they cannot, on the grounds that coercion requires limiting a person’s set of options, and “even if an offer is made to someone facing bleak options, and even if it adds nothing good, an offer does not make the existing set of options worse.”379 Robert Nozick offered this kind of conception of coercion in 1969, and the position that offers cannot be coercive has been common ever since.380 Others, however, hold that the relevant distinction is not between threats and offers, but between those proposals which can genuinely be refused and those which cannot. When the line is drawn in this way, offers can indeed be coercive. On Onora O’Neill’s formulation, an unrefusable offer “is one that a particular victim cannot refuse without deep damage to sense of self or identity.”381 The man who offers to save the life of a mother’s critically ill son if she sleeps with him makes her an unrefusable offer:

---

either she effectively sells her body, or else she allows her son to die when she could have saved him. In either case, her sense of self is damaged, perhaps beyond repair.

In the end, little hangs on whether or not offers can strictly be coercive. Many of those who would refuse to use the term “coercion” for O’Neill’s examples nevertheless recognize that there is something deeply and relatedly problematic about them. David Zimmerman, who is one of the first to consider the possibility of coercive offers, holds that offers like the sex-for-medicine offer above are certainly exploitative even if they cannot strictly be coercive.\textsuperscript{382} And Benjamin Sachs makes an even thinner distinction in a recent paper on coercion: he holds that an offer cannot be an instance of coercion, but can nevertheless be coercive.\textsuperscript{383} Indeed, we can go even further afield, and call offers of O’Neill’s sort “desperate choices”\textsuperscript{384} or instances of “undue influence” instead.\textsuperscript{385} The term coercion is attractive for its rhetorical value, since it provides a graphic reminder of the wrongness of forcing a person into a position where she must choose an option that will harm her in some way. But while I continue to call such instances cases of coercion, they could easily be referred to by any of these other terms as long as the seriousness of the problem is remembered.

If offers can be coercive in this way, then the kinds of incentives that I propose are prime candidates for coercive offers, at least when they are used in developing world contexts. Lisa Fuller does not deal directly with this issue, but provides a useful and closely related discussion of the risk that liberal international NGOs run of making coercive offers to members of “traditional” communities. Fuller makes important points, but her discussion fails to take into account the interests of children. Accordingly, I review her discussion in order both to adopt some of its elements and to highlight the changes that must be made to her account.

I share Fuller’s diagnosis of the problem: in cases where benefits are made available only to those who make some value-laden change in their behavior, “some community members may be very badly off and so may feel that they have little choice


\textsuperscript{383} Benjamin Sachs (2011) “Why Coercion is Wrong When It’s Wrong” \textit{Australasian Journal of Philosophy}.


but to compromise their values in order to obtain the material benefits they need.  

Fuller also identifies an important set of background conditions which NGOs must take into account when they consider the introduction of value-laden services and goods into a community:

(1) the power differential between beneficiaries and INGOs; (2) the significance of a given cultural practice to (at least some of) their potential beneficiaries; and (3) the non-ideal political context in which INGOs must do their work.

The power differential matters because those with more power have the ability to set terms that may be either favorable or unfavorable to the less powerful party, and the less powerful party will often have no choice but to take them – whatever they may be. The responsibility to set terms to which the less powerful party could reasonably agree, then, lies with the more powerful party. The other two elements matter because of the power differential: when a less powerful party places such significance in a practice that it (at least in part) constitutes her identity, she cannot reasonably give it up for the sake of some instrumental good; but when that party lives in precarious and unjust circumstances, she may effectively have no choice but to give up that part of herself if doing so is the only way to access instrumental goods necessary for her survival.

The kinds of incentives that I propose run the risk of forcing persons into this kind of impossible – if not necessarily strictly coercive – choice. Incentives were to be preferred as a means of response to adaptive preferences because they were non-coercive: people could simply refuse to take them up if they had a strong principled aversion to the promoted behavior. But in the developing world, where the material goods offered as incentive might go some significant way towards providing persons with an otherwise-inaccessible minimally decent life, it may not make sense to think of a necessary good as one that can be outweighed by principle. When their need is great enough, even those with the strongest commitments to principle may feel that they have no choice but to violate those principles in order to access the incentive offered. In such cases, the use of incentives will no longer be compatible with showing persons primary recognition respect, since it will run the risk of pressuring them into living their lives in a way that conflicts with their convictions.


387 Ibid, p. 222.
But these considerations, while deeply important, are not the only ones involved. While Fuller rightly argues that it is morally problematic to place persons in this sort of impossible choice situation, she makes several assumptions and omissions that cannot be made if the interests of children are to be given appropriate weight. First and most obviously, she fails to recognize that children who will grow up in a community and do not already place deep significance in its practices also have an interest in whether or not those practices will be preserved — but have no voice in deciding whether they will be. Although Fuller recognizes that persons may invest significance in unjust norms, she holds that those persons themselves must decide whether or not these norms should be preserved.\(^{388}\) I do not deny this. But noting that "it does not seem appropriate to expect aid recipients to be able to shed all of their ingrained socialization under desperate circumstances", Fuller goes on to conclude that it "is no good offering those services that recipients would value if they were living under different institutions or subscribed to a different set of norms."\(^{389}\) But this does not follow. It may not make sense to offer such alternatives for adults' own sake, but it certainly makes sense to make those alternatives available for the sake of children who, through their offering, may become better placed to demand different norms and different institutions in the future.

Secondly, Fuller seems to hold that in order for a counter-normative value-laden practice to be justifiably introduced into a community, it must be one that most people would endorse.\(^{390}\) Certainly, adult members of a community have a legitimate claim to being a part of the processes by which decisions in that community are made. But that legitimate claim does not preclude some resources being justifiably used to protect the (present or future) capacity of a minority to meaningfully engage in that decision-making procedure, even against the will of "most of the people." And this is precisely what I propose when I argue for the use of incentives to ensure that children growing up in a community have the live alternative options that allow them to reflectively endorse or challenge the dominant norms that they find around them.

The challenge is to find a way of constraining the use of incentives so that they can be used to avoid the short-comings of Fuller’s account, while also recognizing the

\(^{388}\) Ibid., n. 4; p. 218.
\(^{389}\) Ibid. p. 224-225.
\(^{390}\) See, for instance, p. 228 where she discusses community members as if they made up one fairly homogeneous "side" in a negotiation over cultural change. (Note that she does go on to note that communities are internally diverse.) See also n. 49 where she notes that extreme views can often be put to the side on the grounds that "usually not everyone in the community holds these views, [and] it is likely that some agreement can be reached that satisfies most people." Italics mine.
important points that she makes. To do this, I propose that we think of preferences as forming a “web”.

A solution: a “web” of preference

W. V. O. Quine holds that our beliefs form a “web”. In his words, “our statements about the external world face the tribunal of sense experience not only individually, but as a corporate body”, so that “revaluation of some statements entails revaluation of others, because of their logical interconnections.” That is to say, our beliefs do not exist in isolation from one another, but rather hang together in such a way that changing one of them will often require changing another. If I believe both that my friend is in town and that it is she who is knocking at my door, then I will be compelled to change both beliefs if I come across strong evidence that she is not in fact in town.

Preferences fit into this web. Some preferences involve desires that are less rational than other sorts of belief, and it is not clear that persons demand the same consistency of preference that they do of other kinds of belief. Accordingly, changes in one preference may not lead as directly to changes in others as in the case of the evidence-based beliefs with which Quine is primarily concerned. Yet changes in one preference certainly can lead us to become more open to changes in those near-by. For instance, if I had previously been averse to soccer, but one day tried it and enjoyed it, this may cause me to revisit my aversion to other team sports. Similarly, if I previously doubted that I would make a good politician on the grounds of my gender, being exposed to successful female politicians might also encourage me to doubt my

392 Quine (1980) A Logical Point of View, p. 41.
393 Ibid., p. 42.
394 I leave aside here broader questions about the relationship between beliefs and desires, and make only the weak claim that desires are in some way – if perhaps only in practice – less amenable to rational persuasion than other kinds of beliefs. For a cognitivist account of desire (and emotion more broadly), see Martha Nussbaum (2001) Upheavals of Thought: The Intelligence of Emotions, Cambridge: Cambridge University Press. For a non-cognitivist account, see Michael Brady (2009) “The Irrationality of Recalcitrant Emotions” Philosophical Studies, 145(3).
396 This idea is based in part on Gerry Mackie’s discussion of the network structure of attitude, which he draws upon to explain how deliberation can fail to lead people to change their minds directly and immediately, but may still lead them to do so indirectly and incrementally over time. Gerry Mackie (2006) “Does Democratic Deliberation Change Minds?” Politics, Philosophy & Economics, 5(3).
previous skepticism of women in executive positions in business. These changes may function as they do in the case of evidence-based beliefs, when transitivity requires a change in one area to lead to a change in another; but such changes may also occur when a change in one preference simply makes one willing to consider new possibilities in other areas where change is not strictly required, but where alternatives had previously seemed too implausible to consider.

Thinking of preferences as forming a web makes it easier to take on board both the strong points and the short-comings of Fuller’s discussion of coercive offers in impoverished circumstances. While near-by preferences are linked, so that a change in one leads to a change in another, near-by preferences need not be equally principled or equally strongly held. For instance, a young woman who feels strongly that she would have an abortion in the early stages of her pregnancy may be unsure what she would do if she had no opportunity to terminate the pregnancy before it reached its later stages. But changes in these less principled preferences can lead to changes in more principled preferences as well – the young woman who struggles with her decision to have a late term abortion may suddenly question the stronger stand that she would have taken in the earlier case.

If preferences are conceived of in this sort of network, then incentives can be used to target a preference for change either directly or indirectly. For instance, if the aim is to encourage women to hold political office, they could be directly targeted by providing an income for women who choose to campaign. They could be indirectly targeted by providing women with an incentive for taking part in a related behavior such as joining a local council to make decisions about issues affecting only women or children. This dual-level targeting can help to solve the problems raised in my discussion of Fuller. Persons may indeed be forced to make desperate choices if the only way of accessing necessary resources is to take advantage of incentives that require them to violate their commitments – but if their near-by preferences are less principled, then incentives can instead be aimed at those preferences whose violation is less costly. This prevents adults from being forced to choose between their commitments and their basic needs, but it also allows for the possibility that changes in less principled commitments will bring people to reconsider their more principled

397 This seems a plausible aim for the use of incentive, since women are far more likely to stand for office when they have already seen other women occupy the same position. See Lori Beaman et al. (2012) “Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India” Science, 335(6068); Lori Beaman et al. (2009) “Powerful Women: Does Exposure Reduce Bias?” The Quarterly Journal of Economics, 124(4).
commitments, so that they may freely decide to change them. In other words, it prevents people from feeling forced into making desperate choices, but it also introduces related alternatives that can help persons to more freely and thoughtfully reconsider what it was that made those choices desperate in the first place. This solution recognizes both the interests that adults have in their current commitments, and the interest that children have in being exposed to new possibilities as they develop their own commitments.

If directly-targeted incentives work better than indirectly-targeted incentives (which, again, will be an empirical question requiring further investigation by social scientists), then they are still to be preferred where they are permissible. But in order to be permissible, they must be consistent with showing persons primary recognition respect. Determining whether directly-targeted incentives are permissible will not be an entirely straight-forward task, but the following three considerations provide a good starting point for determining whether they are:

1. **Community members generally have a sufficient level of resources to achieve a minimally decent life.** When this condition does not hold, it will be more likely that adults will be forced to make a desperate choice. In such cases, incentives should be targeted at preferences indirectly.

2. **The preferences in question are more a function of structure than principle.** For instance, Brooke Ackerly offers an example of women who forwent their own education not because they were opposed to being educated, but because they thought that educating their sons would provide a better financial return than educating themselves. In such cases, it is more permissible to target incentives directly at preferences, because it is circumstances rather than deep values that keep persons from taking up the behavior in question.

3. **There are at least some members of the community who already want to take advantage of the option to which the incentive is attached.** Even if conditions 1 and 2 do not hold, it can still be permissible to target preferences directly in cases where at least some persons in the community would already prefer to violate that norm. Indeed, this is probably the preferable circumstance for the use of incentive. These counter-stereotypic individuals will have a better idea of the kinds of non-normative behaviors that can best be undertaken, and they will own those behaviors in ways that others encouraged only by the material benefits of incentives may not. However, incentives should only be directly targeted in such cases if other means of material support are also made available to the community more broadly. In this way, those who want to take up the incentive can, but those who do not need not decide between their commitments and material necessities.

---

Conclusion

One final caveat. I have been speaking as if coercion, deliberation, and incentive are clear and distinct categories of potential response to adaptive preferences. As the discussion should suggest, however, this will not always be true. Often the categories will blend together and there will be elements of more than one in whatever method of response is ultimately adopted. Coercion should be avoided whenever possible, but no incentive-based method of response to adaptive preferences could be either fully justified or usefully implemented without a strong deliberative component. As I hope that I have succeeded in showing in the previous chapters, however, many deliberative methods of response to adaptive preferences will fail to take seriously the needs of children if they rely only on deliberation – and accordingly, they also will be unjustifiable. Pulling deliberation and incentive apart in practice may be difficult. The use of incentive may cause a group of persons to take up a certain behavior and therefore come to see it (and the values related to it) as live. But things may also work in the other direction. Through a process of persuasion, persons may come to realize – or simply be provided with an easy avenue for expressing – that they would like the set of socially available goods to be different. In this case, resources devoted to facilitating persons in taking up new patterns of behavior would be less plausibly described as incentives encouraging persons to take up a course of behavior that they would not otherwise have adopted, and more plausibly understood as the substantive means that allowed persons to take up a new course of action that was previously desired but was not previously substantively available.

My discussion of incentives is supposed to call for an increased emphasis on the role of incentives and the interests of children in response to adaptive preferences, not an exclusive one. Where possible, deliberation is to be favored, both pragmatically (because of the difficulty of determining the interests of adults in our world of limited epistemic access to the good of individuals) and for more principled reasons (since deliberation increases the agency of persons and the extent to which they may be the architects of their own lives). But as I have shown, there are good moral reasons to recognize cases in which deliberation fails to do justice to the needs of children – and as I have also shown, this kind of recognition need not justify steamrolling the rights and interests of adults. I have outlined a way in which the rights and interests of children and adults can both be taken seriously.

198
The argument in this chapter has shown how the remaining considerations developed in the first half of the thesis can be taken into account. It helps with both Consideration 2 and Consideration 7:

**Consideration 2: Variety of preferences.** An account of adaptive preferences should be able to respond to preferences of all types, from those that are deeply and sincerely held to those that represent compromises with unfavorable realities.

**Consideration 7: Efficacy.** An account of adaptive preferences should be conducive to providing solutions that those affected will themselves endorse, since only solutions that affected persons endorse have a significant likelihood of being effective.

Consideration 2 is addressed because incentives can be used to change both those preferences that are deeply and sincerely held, and those preferences that are more a function of structure than of principle. Consideration 7 is addressed because incentives non-coercively change the reasons for action within a situation. Accordingly, they can be taken up only by those who endorse the change in their own behavior, and need not be pushed on those who would resist and reject such change. But most importantly, the account in this chapter takes seriously Consideration 10:

**Consideration 10: Children.** An account of adaptive preferences must be sensitive both to the adaptive preferences that adults already have and the adaptive preferences that children might acquire. Currently held adaptive preferences should be treated differently from those that might yet form.

By arguing for the use of incentives, I propose a method of response to adaptive preferences that works within our highly non-ideal world to show respect both for the preferences that adults already have and may have an interest in maintaining, and for the interest that children have in developing from the beginning preferences that will leave them best able to direct the course of their own lives and transform their world into the world that they want it to be.
Conclusion

Adaptive preferences fall into the set of philosophical subjects in which the importance of practical problems far outstrips the importance of theoretical problems. As I noted in the Introduction, without the political project associated with the concept of adaptive preferences, far fewer theorists would bother to talk about the concept at all. This political project clearly cries out for practical and concrete real-world solutions, and I admit that the work that I have done in this thesis goes only a little way towards providing them. Nevertheless, I hope that the work done here will be of use for the political project. There are several reasons for optimism.

First, I have offered an explicit framework for understanding the phenomenon of adaptive preferences that clearly explains which preferences ought to count as adaptive in theory. This framework explains what is wrong with adaptive preferences and targets the areas in which social changes must be made in order for such preferences to be overcome. As such, it provides a valuable tool for the political project.

Second, the framework that I have offered proves that use of the concept of adaptive preferences, when correctly conceived, is compatible with showing two centrally important forms of respect for persons judged to have adaptive preferences. This work is crucial for the political project, since any formulation of the concept that failed to show persons these forms of respect should rightly be rejected by theorists and practitioners alike. My thesis makes three major advances in this area. By using substantive content indirectly to determine the autonomy of preferences, my theory of adaptive preferences recognizes both that persons’ considered and informed convictions should be the yardstick by which their own good is measured, and that obstacles to reasoning clearly about one’s own good cannot be guarded against by procedural means alone. Instead, opportunities with valuable content must be available in order for persons to have the tools to make informed judgments about their own interests. In this way, my theory of adaptive preferences shows respect for persons by acknowledging that any conception of the good could be autonomously endorsed, without ignoring the social circumstances that can impinge upon human beings’ ability to exercise their unique powers of moral reasoning. Relatedly, by defining adaptive preferences as a revisable form of autonomy deficit, it shows what it is that is wrong with adaptive preferences without holding persons hostage to the way in which their
preferences were formed. Since initially adaptive preferences can be reflected upon under the right circumstances and thereafter retained autonomously, my theory treats persons' own considered convictions, and not merely the history of their preferences, as the criteria by which they are judged to be worthy of suspicion. In this way, my theory of adaptive preferences respects persons as competent moral reasoners while also recognizing that persons' circumstances can unjustly impinge upon their ability to exercise those powers. And finally, by placing autonomy at the center of my theory of adaptive preferences, I offer a strong reason to develop social circumstances and institutions that increase the abilities of all persons to expand their powers of autonomy. In this way, my theory respects persons by recognizing the interest that they have in living their lives in accordance with convictions that are genuinely their own, and demands that society provide them with the tools to do so.

Third, and perhaps most importantly, my theory of adaptive preferences expands the discussion of adaptive preferences to include the needs of children. Instead of focusing inadvertently on a toy world in which only the interests of adults matter, my theory demands that any discussion of adaptive preferences do justice to the world in which we actually live - that is, a world comprised of both children and adults, where each group has different, but equally important needs. My theory takes seriously the fact that the interests of children and adults in this area can conflict, and recognizes the difficulty of doing justice to both groups at once. Since my thesis represents the first in-depth philosophical contribution to this topic, much work remains to be done in this area. I hope, however, to have laid the groundwork for a promising and fruitful further conversation.

In part, the work that I have done on children and adaptive preferences requires further attention because many of the most important differences between children and adults arise at the levels of identification of and response to adaptive preferences. Identification and response must always involve actual preferences in actual, concrete circumstances, and concrete circumstances will always differ from one another in important ways that cannot be accessed from the armchair. Accordingly, doing justice to both children and adults when it comes to the identification of and response to adaptive preferences is not a task that can be undertaken by philosophers alone. Instead, philosophers, social scientists, and practitioners must all be involved in a sustained project of both applying and revealing the limits of the theory that I have begun to lay out. This is as it should be. Moving beyond theory is exactly what the political project requires.


Anca Gheaus (2011) "Arguments for Nonparental Care for Children" Social Theory and Practice, 37(3).


Elizabeth Harman (2009) "‘I'll Be Glad I Did It’ Reasoning" Philosophical Perspectives, 23.


Steven V. Mazie (2005) "Consenting Adults? Amish Rumspringa and the Quandary of Exit in Liberalism" Perspectives on Politics, 3(4).


National Center for Education Statistics (nces.ed.gov).

The New Yorker (January 2, 2012) "Drug Test".


Brian A. Nosek (2007) "Implicit-Explicit Relations" Current Directions in Psychological Science, 16(2).


