

AGNES CALLARD

Aspiration



*the agency of
becoming*

OXFORD
UNIVERSITY PRESS

This book is dedicated to my teacher Amy Kass,
who made her classroom a theater of aspiration.

Table of Contents

Introduction

- I. Overview: Reasoning Toward Value, p. 5
- II. Outline of Chapters, p. 8
- III. A Case Study: Alcibiades, p. 11
- IV. Self and Value, p. 20

Part I Practical Rationality

- CHAPTER 1: Decision Theory and Transformative Choice, p. 24
 - I. Ullmann-Margalit on Old Person vs. New Person, p. 25
 - II. Paul on Becoming a Vampire, p. 26
 - III. Self-Cultivation vs. Transformation, p. 28
 - IV. Privileging Second-order Preferences? p. 29
 - V. Reasoning Toward Desire: Rational Self-Transformation, p. 31
 - VI. Deciding to Become Someone New? p. 38
- CHAPTER 2: Proleptic Reasons, p. 41
 - I. Large-Scale Transformative Pursuits, p. 44
 - II. Alternatives to Proleptic Reasons? p. 45
 - (a) Vague Reasons
 - (b) Self-Management Reasons
 - (c) Testimonial Reasons
 - (d) Reasons of Competition
 - (e) Reasons of Pretense
 - (f) Approximating Reasons
 - III. A Dilemma: Flailing vs. Deepening, p. 51
 - III. Internal Reasons, p. 53

Part II Moral Psychology

- CHAPTER 3: Intrinsic and Extrinsic Conflict, p. 61
 - I. Framing Intrinsic Conflict, p. 62
 - II. Valuing: A Hybrid Account, p. 63
 - III. The Possibility of Intrinsic Conflict, p. 66
 - IV. Deliberative Irresolvability, p. 68
 - V. Intrinsic Conflict and Ethical Theory, p. 71
 - VI. Two Suggestions for Resolving Intrinsic Conflicts, p. 73
 - (a) By Identification (Frankfurt)
 - (b) By Aspiration
- CHAPTER 4: Akrasia, p. 81
 - I. Acting on a weaker reason? p. 83

- II. Akritic Conflicts Are Intrinsic, p. 85
- III. Which Reason is the Akritic One? p. 89
- IV. Akrasia and Aspiration, p. 94

Part III Moral Responsibility

CHAPTER 5: The Problem of Self-Creation, p. 98

- I. Self-Endorsement, p. 100
- II. Self-Cultivation, p. 102
- III. Neurath's Boat, p. 106
- IV. Grounding Relations, p. 109
- V. Normative Grounding and Self-Creation, p. 111
 - (a) Priority of Created Self
 - (b) Self-Creation by Aspiration
- VI. Teleology and Agency, p. 116
 - (a) Normative Grounding and Backwards Causation
 - (b) Two Conceptions of Agency
 - (c) Aspirational Agency vs. Self-Standing Action

CHAPTER 6: Implications of the Aspirational Account of Self-Creation, p. 123

- I. Aristotle on Habituation, p. 123
- II. Aspiration vs. Dialectical Desire, p. 135
- III. Aspiration vs. Ambition, p. 138
- IV. Responsibility for Self: An Asymmetrical Account, p. 140
- V. The Aspiring Gangster, p. 142
- VI. Responsibility for Character: A Test Case, p. 146

Conclusion, p. 154

Bibliography, p. 163

* Introduction

I. Overview: Reasoning Toward Value

We can all think back to a time when we were substantially different people, value-wise, from the people we are now. There was a time when we were not even aware of the existence of some of the people, activities, institutions, and ideologies that now figure centrally in our lives. Maybe we had different political views or no political views at all; we used to be religious, or used not to be; maybe we now feel deep ties to a place which is spatially, culturally, and linguistically far from where we grew up; maybe we find our interests and concerns resembling those of our parents more than we ever thought they would. We care about many things that we once did not care about. How did that change come about?

In accounting for the genesis of our new values, we often have occasion to mention the effects on us of forces outside our control: a fortuitous coincidence, an influential mentor, an inspiring locale, a tragic loss, a bitter betrayal, a domineering parent, the emergence of an innate facility, the process of getting older. But these kinds of factors will generally only make up some of the story: a mentor cannot implant a love of music; the betrayal cannot, of itself, create a devotion to independence; coincidence cannot produce love; being in a culinary mecca cannot make one into a chef; talents do not develop themselves. There is no doubt that our parents, friends and romantic partners influence us deeply, but they do not fashion us. We have a hand in answering the question as to what things in the world are important to us, and our answers need not be, and typically are not, arbitrary or random. But agency, as distinct from mere behavior, is marked by practical rationality. Insofar as becoming someone is something someone *does*, and not merely something that happens to her, she must have access to *reasons* to become the person she will be. Giving a philosophical account of how it is possible for value-acquisition to be a form of practically rational agency is the project of this book.

This project faces a difficulty: people do not seem to be able to *choose* or *decide* to have different values. A decision or choice is an act of the will that prefigures, accompanies, or is a constitutive part of some single action. The transition from indifference to love cannot typically be effected by way of doing any one thing: no matter the strength of my will, it does not seem that I can muscle myself into suddenly caring. To be sure, the path to valuing sometimes includes momentary expressions of commitment: the moment when you say “I do”, or sign the adoption papers, or buy the one-way plane ticket to a foreign country. But these moments are themselves only part of the story, punctuating a longer process. Coming to value something tends to represent a deep change in how one sees and feels and thinks. Acquiring a new value often alters the structure of one’s priorities by demoting or even displacing something one valued before. Such changes take time, over the course of which one has done many different things in the service of value-appreciation. The later actions are shaped by the small changes that the earlier ones have engendered in such a way as to allow someone to slowly develop new priorities, concerns and attachments. The process as a whole exemplifies a distinctive form of practical rationality, one not structured by a single moment of intention or decision at its inception; the rationality of the agent I seek to describe changes and indeed solidifies over time, as the agent becomes increasingly able to respond to the reasons for action associated with her new values.

We have a rich vocabulary for differentiating the many forms that positive practical orientations can take: in addition to valuing, we speak of desiring, wanting, loving, approving of, being attracted to, caring about or for, endorsing, preferring, being identified with, seeing as valuable etc. I will set aside any differences that such terms are sometimes taken to mark, in order to focus on the rational process, to the extent that there is one, by which we arrive at any of these “pro-attitudes,” as Davidson calls them (1980¹, p.4). I myself avoid the term “pro-attitude” only because I fear that such a term of art is liable to alienate us from our own intuitions about the process we are describing. For this reason, I prefer to speak, as we colloquially do, of agents coming to value, desire, want, etc.; nothing should be taken to hang on the choice of one such term over another.

Grasping new values is hard for us because, to paraphrase Augustine¹, our hands are already full. Without denying that parents and teachers may play an important role in such a process, we might nonetheless characterize it as one in which one habituates or educates oneself. It is a mark of being old enough to engage in such an activity that the one already has interests, concerns, and projects that can serve as obstacles in the task of acquiring new ones. Gaining a value often means devoting to it some of the time and effort one was previously devoting elsewhere. Sometimes one’s new value requires complete divestment from an old value, for instance when a former pleasure-seeker turns herself toward asceticism. Even in cases where our old value-outlook does not specifically contradict our new one, we often experience the effort of coming to apprehend value as a struggle with ourselves. Leisurely self-contentment is ruled out for someone who sees herself as being in a defective valuational condition. Grasping new values is hard work.

The name I will give to the rational process by which we work to care about (or love, or value, or desire...) something new is “aspiration.” Aspiration, as I understand it, is the distinctive form of agency directed at the acquisition of values. Though we do not typically come to value simply by deciding to, it is nonetheless true that coming to value can be something the agent does. The explanation of how we come to value, or to see-as-valuable, so many of the things that we once did not is that we work to achieve this result. The aspirant sees that she does not have the values that she would like to have, and therefore seeks to move herself toward a better valuational condition. She senses that there is more out there to value than she currently values, and she strives to come to see what she cannot yet get fully into view.

The work of aspiration includes, but is by no means limited to, the mental work of thinking, imagining, and reasoning. If a callow youth gets an inkling of the value of classical music or painting or wine, and wants to come to appreciate these values more fully, he must listen to music or visit museums or drink wine. Let me offer a few more examples, some of which may strike the reader as more familiar than others. If one aspires to be a doctor, one goes to medical school. If one aspires to be more attuned to values of healthy living, one might become a member of a gym and transition one’s eating habits toward eating more vegetables. If one seeks to appreciate some person, one might invite him for coffee. If one aspires to be religious, one might spend more time at one’s church or synagogue or mosque—or, in another kind of case, one might deliberately stay away from those places in the effort to (re)connect with God on

¹“God wishes to give us something, but cannot, because he sees that our hands are already full,” cited in Schillebeeckx p.242.

one's own terms. If one's goal is to value civic engagement, one might explore community activism. We aspire by doing things, and the things we do change us so that we are able to do the same things, or things of that kind, better and better. In the beginning, we sometimes feel as though we are pretending, acting or otherwise alienated from our own activity. We may see the new value as something we are trying out or trying on rather than fully, committally engaged with. We may rely heavily on mentors whom we are trying to imitate or competitors whom we are trying to best. As time goes on, however, the fact (if it is a fact) that we are still at it is usually a sign that we find ourselves progressively more able to see, on our own, the value that we could barely apprehend at first. This is how we work our way into caring about the many things that we, having done that work, care about.

The English word "aspiration" is a good, if not a perfect, label for the concept I aim to explicate. Since I use the word to describe the process of rational value-acquisition, I end up emphasizing certain of the ordinary language features of the word and de-emphasizing others. For instance, we often speak of someone's aspiring to some career, as I did a moment ago, when describing an aspiring doctor. In this kind of context, we may think that such a person's primary hope is to acquire the skills and qualifications that further enable her to secure extrinsic rewards such as status, money, or parental love. The aspirant, as I use the word, never aims exclusively at any of these things. To be sure, she wants to go to medical school, to pass her exams, to succeed in her residency, to gain a position at an excellent hospital. Perhaps she even wants to please her parents. But her desire for all these things is a secondary manifestation of what she really wants, which is access to the distinctive value of helping people with their medical problems. Though she takes herself, pre-medical school, to have some conception of what that is and why it is valuable, she (knows that) she does not really know what it is like before engaging in the work whose value she wants to access.

A medical student whose final target was money, the approval of her parents, or social status would not count as an aspirant in my sense; I discuss this distinct phenomenon, which I call "ambition," in chapter six. The ambitious medical student is not seeking to acquire a value: she takes herself to have full access, even before entering medical school, to the value of having money, the approval of her parents, or social status. She does not hope that medical school and residency will teach her the value of these things. She hopes only that it will help her satisfy the values she already has. She has too much access to the value in question to count as aspiring toward it. More generally, the word 'aspiration' is sometimes used interchangeably with having a hope or wish or long-term goal to bring some result about. These agents will not count as aspirants in my sense, unless the sought-after end is one whose value those agents are also seeking to learn.

I will also restrict the term 'aspirant' at the other end of the spectrum, by withholding it from people who have *too little* antecedent access to any value that they might acquire. It is not a stretch of the English word 'aspirant' to describe a young adult who sets out to Europe to 'find herself' as an aspirant. She won't count as an aspirant in my sense, however, unless there is something more specific she is trying to find². Adventures are not typically aspirational, and a

² In my paper "Liberal Education and the Possibility of Valuational Progress," I identify a form of aspiration that may be an exception to this rule. Colleges and Universities provide aspirants with the kind of support that makes it

sign of this is that they rarely feel like work³. The aspirant is trying to change herself in some particular dimension; she is not merely open to changes that might come. She grasps, however dimly, a target with reference to which she guides herself.

It is not always easy to determine how much of an antecedent grasp of value someone has, nor to ascertain how much of a grasp someone would have to have in order to count as an aspirant. I won't offer any guidance for assessing borderline cases, though I will discuss why this is a difficult problem, and why such assessments may presuppose specialized knowledge of the field in question (see ch. 2 p. 83). My point here is only to note that my use of the word "aspirant" is philosophically charged in such a way as to pick out all and only the cases in which the project of becoming someone is also the process of appreciating the values distinctive of becoming that kind of person.

Aspiration is rational, purposive value acquisition. In the second part of this introduction, I offer a case study of an aspirant taken from Plato's *Symposium*. Alcibiades' closing speech gives us access to what it feels like to struggle to be better than one is, and Plato's presentation of that speech makes it possible for us to assess the rationality of Alcibiades' attempts at value-acquisition. My discussion of Alcibiades presupposes that there is such a thing as a rational pursuit of one's own fundamental values; this is the claim which I spend the rest of the book defending. Before turning to Alcibiades, I now briefly outline the structure of that argument.

II. Outline of Chapters

Throughout this book, I describe myself as "giving an account" or "presenting a theory" of aspiration. As this outline will make clear, the work of this book is somewhat more rudimentary than those descriptions might suggest. The topic of aspiration lies at the crossroads of three sub-areas of ethics: the theory of practical rationality, the theory of moral psychology, and the theory of moral responsibility. In each of the three areas, the concept of aspiration emerges as a problematic one—it is difficult to see how aspiration can be rational, how it can be psychologically real, and how it is possible for someone, via aspiration, to 'create himself.' I aim to identify those elements of the received framework in each field that foreclose the possibility of aspiration, and to propose emendations that would accommodate it. This book describes what an aspiration-friendly theory of rationality, moral psychology, and moral responsibility would look like. In addition, by way of motivating my emendations, I explain the payoffs available, in each area, for making the required changes. The theory of aspiration must begin somewhat earlier than the theory of a phenomenon that does not require such emendations; it must begin with an explanation of how it is so much as conceivable that human beings aspire.

Rationality

If aspiration is to be an exercise of human agency, aspirants must be responding to practical reasons of some kind. Behavior qualifies as agency insofar as it exhibits the distinctive intelligibility of being a response to reasons. "I do what happens," as Anscombe said (p. 52),

possible for a person to aspire even if she possesses only an aspirational goal as vague and schematic as that of "becoming someone" or "learning how to think" or "self-discovery."

³ I develop this idea by contrasting aspiration with Talbot Brewer's "dialectical activities" in ch. 6 part II.

but only when what happens happens for some reason. There are, however, problems in identifying the reason on which the aspirant acts. In chapter one, I explore two recent attempts to account for aspirational activity within the framework of decision theory. If that project fails, then it might seem—as one of the authors in fact concludes—that the process of substantive value change simply cannot be rational. I argue instead that it has a distinctive rational form that is not the rationality of deliberation, calculation, preference or decision.

In chapter two, I discuss the special practical reasons peculiar to aspirants, which I call “proleptic reasons.” If someone takes a music class in order to come to appreciate music, her behavior does not serve a current end of hers in the same way it would if she got a cheeseburger because she was hungry. In the second case, she already has a desire for food; in the first, she is trying to have a desire for music. The reasons of aspirants are not, to use Bernard Williams’ term, ‘internal reasons’ to which an agent can expect complete access if she deliberates correctly from her current motivational condition. Nor are they the purely ‘external’ reasons sometimes moralistically ascribed to agents with reprehensible motivational makeups. For instance, we might be inclined to say, of a rich person who (let us suppose) we know to be constitutionally incapable of becoming motivated to help the poor, “he nevertheless has reason to donate money to charity.” The aspirant’s reasons are, by contrast, ones she is in a position to appreciate—though not fully. The theorist of aspiration identifies a kind of reason that is not fully ‘external’ to an agent’s motivational condition, but that also fails to be fully accessible to her at the outset of her project. These ‘proleptic’ reasons are the reasons grounding the rationality of any process of substantive value change.

Moral Psychology

There is a characteristically aspirational form of angst. In order to bring out what is distinctive about the aspirant’s inner strife, it will be helpful to contrast it with two well-recognized sources of psychological conflict: that of a hard or tragic choice, and that of a recalcitrant or rejected motive. In the case of a hard choice, an agent may find it difficult to decide which of two options to choose. Perhaps the values are incommensurable, or perhaps she simply does not want to give up on either: even getting what is better overall involves a substantial loss. Such an agent feels pained at the loss of whichever option she lets go of.

This kind of pain is quite different from that of a recalcitrant or rejected motive. Consider the unwilling addict. She is moved by forces that she views as in some way external to her will. Her motivation to take drugs is not accompanied by the corresponding evaluative judgment to the effect that she ought to take the drugs. The case of the addict is typically taken to be an instance of the more general phenomenon of alienation from affective conditions such as pathological fears or bouts of uncontrollable rage. Such a person might feel that there is nothing to be said for feeding her addiction (or fueling her rage or accommodating her phobias), but she is moved to do so by inner drives she cannot control.

The phenomenon of aspiration opens up a third way of being torn. Though she looks forward to a time when she will no longer find operas boring, the aspiring opera lover does not currently find her boredom external or alien. It is all too clear that the indifference she feels really is *hers*, which is to say, it represents a point of view that she identifies as authentically her own. This is exactly why she (sees that she) needs to work to see things differently.

Nor does the aspiring music lover find the choice between love of and indifference to music to be a ‘hard choice.’ She is oriented toward the one condition, and away from the other, in such a way as to make the decision an easy one. She does not feel uncertain, nor does she feel that by coming to love music she is choosing between two things, her love and her indifference, both of which are really important to her. Nonetheless coming to love music can be difficult. Someone who is working at it will often feel torn. I describe this form of conflict, which I call ‘intrinsic,’ in chapter three.

Aspirants are not the only kinds of agents who experience intrinsic conflict. In the fourth chapter, I bring out the special character of the aspirant’s intrinsic conflict by contrasting it with the intrinsic conflict experienced by the akratic. I begin by articulating a standing problem with analyses of akrasia since Davidson: they force us to choose between saying that the akratic agent did not (really) know that she should have done otherwise, or saying that she did not act willingly. Effectively, the agent is depicted either as not having fully decided the issue between her two options—it was a hard choice!—or as having been overwhelmed by a form of motivation that is alien or external to her. I argue that a better account of akrasia is available to those who understand it as an instance of intrinsic conflict. The same modifications to the existing moral psychological structure that are needed to make room for aspiration also offer up a neglected alternative for understanding the nature of the akratic’s conflict.

Akrasia, I argue, is aspiration writ small: when we try to describe a case of aspiration without allowing for the dimension of extended temporality, what we get is a case of akrasia. While aspiration may be philosophically neglected, akrasia cannot claim to be. Those who puzzle over how it is possible to act against one’s better judgment, are, I argue, grasping a tip of the aspirational iceberg. It is worth noting, however, that my analysis of akrasia is, in an important sense, freestanding from the rest of the book. One needn’t accept that akratics are intrinsically conflicted in order to embrace my account of the role of intrinsic conflict in aspiration; rather, the direction of support goes the other way. In applying the framework of intrinsic conflict to the paradoxical phenomenon of akrasia, I aim to illustrate the explanatory power of the theory of aspiration.

Responsibility

In the final section of the book, I argue that the theorist of aspiration offers us an important piece of a solution to a long-standing puzzle as to how we can be responsible for being the kinds of people we are. Much of who we are either is or flows from our values—but we cannot, according to this puzzle, choose our values. For if we are inclined to choose to value something, that must (so goes the dilemma) be because the valuing of that thing already accords with the values we have—which is, of course, why we are so inclined. Otherwise, coming to value it entails breaking radically with our old values, in which case the transition is neither rational nor agential. The new valuation, on this horn of the dilemma, is something that happens to us, rather than something we do.

In chapter 5, I show that we can escape the dilemma by inverting the traditional relation of authority between the creator and the created self. On an aspirational account of self-creation,

the creator does not determine, choose or shape the created self; rather, she looks up to, imitates and seeks to become the created self. The source of normativity lies at the end of the process, rather than at the beginning.

In chapter 6, I develop some ethical implications of this approach to the phenomenon of self-creation. I explain that it offers richer conceptual resources for, first, interpreting Aristotle's account of habituation, and, second, offering an account of someone's moral responsibility for our valuational condition. We are positively responsible—praiseworthy—for the good valuational condition we attain via aspiration. We are negatively responsible—blameworthy—for the culpable failure to aspire to a better condition. The account is asymmetrical because aspiration is a learning process: since one can't learn what is not the case, there is no such thing as 'aspiring to be evil.'

The expression 'finis origine pendet' means, the end hangs from (i.e. depends on) the beginning." It is attributed to the Roman poet Manlius and was adopted as a motto by Philips Exeter Academy. On one interpretation, the phrase asserts that what happens in the early years has a substantial impact, positive or negative, on the later unfolding of that life. Perhaps this is all Manlius meant to say; perhaps he was simply pointing out that beginnings are important. But once we are operating in the educational context into which the founders of Exeter Academy imported the phrase, it seems fair to append to the motto the clarification that the most important beginnings are those that, in an aspirational sense, hang from the end.

III. A Case Study: Alcibiades

As the above overview suggests, the very possibility of aspiration has been neglected in the philosophical literature in rationality, moral psychology and responsibility. The effects of this neglect are visible not in any overt claim that such a thing is *impossible*, but in certain subtle ways in which the space of possibility has been narrowed in each of the three sub-fields discussed above. The theorist of aspiration finds herself pointing to something in the middle of what we might have taken to be an exclusive dichotomy between internal and external reasons; and between 'choosing' a value and being saddled with one; and between the unwilling addict's alien motivation to take the drug and a form of motivation that one wholeheartedly embraces. In the face of the standing assumption that there is a principled distinction between a process of discovery and one of creation, she points out that the aspirant's value-discovery is at the same time her self-creation. She is forced to stretch the existing concepts into a new dimension by pointing to *degrees* to which one sees a reason, or has a value, or inhabits some point of view.

We cannot prove the need for these innovations simply by offering an example of aspiration, since a skeptic can always read any example in a non-aspirational way. Examples have a limited power to defuse a skeptic bent on reductive elimination of the phenomenon (purportedly) exemplified in them; this problem is particularly acute in the case of aspiration, which can be hard to distinguish from related phenomena such as ambition, adventurousness, indoctrination and pretense. Nonetheless, an example will help us lay out what an aspirational account of some phenomenon would look like. And a well-written example can do a bit more. For an author who has taken some care to present the details of an aspirant's psychology in a realistic way affords the theorist of aspiration an opportunity to showcase the interpretative

power of her innovations in the theory of rationality, psychology and ethics. I find such an opportunity in the speech of Alcibiades at the end of Plato's *Symposium*. As a first step in arguing for the possibility of aspiration, I propose to argue that an aspirational account of Alcibiades' rationality, psychology, and ethical status offers us the least contrived reading of Plato's text.

In Plato's *Symposium*, Socrates attends a drinking party, the participants in which take turns offering speeches of praise to the god Eros. As the last speaker is finishing up, a drunken Alcibiades crashes the party. Alcibiades was, at the time, a newly minted general whose good looks, charisma and aristocratic lineage engendered a large base of popular support. Alcibiades' throng of suitors as a teenager is described in Plato's *Alcibiades I*. In the *Symposium* he is a bit older, and what he has lost in boyish appeal—upon entering, he removes a laurel crown from his own head and places it on that of the beautiful young up-and-coming poet, Agathon—he has gained in political influence. Indeed, given his victories in the 416 Olympic games⁴, and his recent appointment as general, it is fair to say that in the *Symposium* Plato catches Alcibiades on the cusp of greatness—a greatness that, as we will have occasion to discuss further below, he was never to attain.

Alcibiades opts not to give a praise of Eros, as the other symposiasts have done, but to praise Socrates instead. Because his encomium issues from Alcibiades' vantage point as Socrates' would-be lover, it is as much a lamentation of Alcibiades' sufferings as it is a catalogue of Socrates' virtues. Alcibiades recounts the practical consequences of the famed 'knowledge of one's own ignorance' with which Socrates infects those around him. He describes his newly discovered need to live a wholly different life, to become a much better person than he is. Unthinkable, that *Alcibiades* could feel inadequate!—at least Alcibiades finds it so. What one hears throughout Alcibiades' speech is the shock of finding *himself, of all people* in this position.

Alcibiades discusses his own aspirational condition at length, because he is so struck by the unlikeliness of it. But even as he describes it, he is *in* it—so his speech offers us access to the phenomenon of aspiration, described from the point of view of the aspirant. He tells us both what aspiration looks like, and what it feels like.

“You know, people hardly ever take a speaker seriously, even if he's the greatest orator; but let anyone—man, woman or child—listen to you or even to a poor account of what you say—and we are all transported, completely possessed. If I were to describe for you all what an extraordinary effect his words have always had on me (I can feel it this moment even as I'm speaking), you might actually suspect that I'm drunk! Still, I swear to you, the moment he starts to speak, I am beside myself: my heart starts leaping in my chest, the tears come streaming down my face, even the frenzied Corybantes seem sane compared to me—and, let me tell you, I am not alone. I have heard Pericles and many other great orators, and I have admired their speeches. But nothing like this ever happened to me: they never upset me so deeply that my very own soul started protesting that my life—my life!—was no better than the most miserable slave's. And yet that is exactly how this Marsyas here at my side makes me feel all the time: he makes it seem that my life isn't worth living! You can't say that isn't true, Socrates. I know very well that you could make me feel that way this very moment if I gave you half a chance. He always traps me, you see, and he makes me admit that my political career is a waste of time, while all that matters is just what I most neglect: my

⁴ Mentioned by Alcibiades himself in Thucydides VI.16.2, where one can also find a sketch of Alcibiades' arrogance and influence that corroborates Plato's.

personal shortcomings, which cry out for the closest attention. So I refuse to listen to him; I stop my ears and tear myself away from him, for, like the Sirens, he could make me stay by his side till I die. Socrates is the only man in the world who has made me feel shame—ah, you didn't think I had it in me, did you? Yes, he makes me feel ashamed: I know perfectly well that I can't prove he's wrong when he tells me what I should do; yet, the moment I leave his side, I go back to my old ways: I cave in to my desire to please the crowd. My whole life has become one constant effort to escape from him and keep away, but when I see him, I feel deeply ashamed, because I'm doing nothing about my way of life, though I have already agreed with him that I should. Sometimes, believe me, I think I would be happier if he were dead. And yet I know that if he dies I'll be even more miserable. I can't live with him, and I can't live without him! What can I do about him? (215d1-216c3⁵)

Alcibiades is torn between being inclined, with Socrates, to “admit that my political career is a waste of time,” and the powerful inclination to pursue that same career. He is sincere⁶ when he acknowledges the importance of attending to the personal shortcomings he goes on to neglect. For even when he is not talking to Socrates, Alcibiades experiences the Socratic point of view as an oppressive presence pouring forth censure onto his way of life. Alcibiades channels Socrates when he castigates his own life as “not worth living” or as no better than that of a slave; and when he says that he “neglects himself while attending to the affairs of Athens,” his language has an authentically Socratic ring. He claims to know that Socrates is right while nonetheless being overcome by the value—honor—that Socrates has taught him to discount⁷.

Alcibiades insists on a vivid and intense access to the experience of being refuted by Socrates, even now (215de, 216a), that is, when he is not being refuted by Socrates. He dismisses his pursuit of honor as something that ‘overcomes’ him or as a vestige of his ‘old ways’. He describes the effects of Socrates’ speech as something he can “still feel even at this moment” – but this cannot quite be right. For when Socrates refutes him, Alcibiades, by his own reckoning, behaves like a Corybant: “I find my heart leaping and my tears gushing forth at the sound of his speech.” (215e) But his heart is not, as he speaks, leaping, nor are tears gushing forth. If Socrates *could* make him feel that his life is not worth living, that can only be because he doesn't currently feel that way. Alcibiades is clearly referring to an experience that both he and others have had at another time, namely, when they were being refuted by Socrates.

Alcibiades feels that he has a grip on what ‘Socratism’ is and what it does to him: “I know very well that you could make me feel that way this very moment if I gave you half a chance.” (216a) This sentence expresses in a wonderfully vivid way Alcibiades’ sense that he both can and cannot make contact with the Socratic experience from a distance. Alcibiades is presently aware of his own weakness, of just what he *would be* experiencing at the hands of Socrates. But if Socrates were presently refuting him, that weakness would feel different—worse.

⁵ Translations of Plato throughout are from Cooper.

⁶ Plato uses his dramatic resources to credentialize Alcibiades’ speech. He has Alcibiades open with the following invitation to Socrates: “if I say anything that’s not true, you can just interrupt, if you want, and correct me; at worst, there’ll be mistakes in my speech, not lies.” (214e) The fact that Alcibiades is then not interrupted suggests that Socrates, at any rate, does not find Alcibiades’ description to be wildly off.

⁷ In the first few paragraphs of this analysis I draw on the corresponding discussion in Callard 2014. In that paper, I analyze the aspirations of Alcibiades in terms of the understanding of *akrasia* Socrates develops in the *Protagoras*, but discussion of Socrates’ account of *akrasia* would take us too far afield.

Alcibiades has some grip on the kinds of things Socrates will say to him, and the ways that his own actions, choices, and desires will look and feel to him when he is talking to Socrates. But they don't quite look or feel in those ways. He doesn't, as he speaks to the assembled company, hear Socrates' actual voice, but a simulacrum of Socrates' voice, one that uses Socratic phrasing but lacks the full Socratic bite. When Socrates actually begins to speak, those accusations will ring much louder. Alcibiades experiences a characteristically aspirational form of torture: he can almost see what it would be like to see things differently, but he doesn't get all the way to seeing them differently.

What Plato brings out so nicely in the speech of Alcibiades is the way in which one's old way of seeing things makes trouble for the new way. Alcibiades' conventional honor-loving values move him to flee from and hate Socrates, even as he struggles to recognize that the honor-loving life is not worth living. He can almost see his values as bankrupt—but he can't quite, because they won't go away.

But is Alcibiades really striving to be different? In this opening speech, it is perhaps not completely clear what aspirational work Alcibiades is doing. For what comes to the forefront of Alcibiades' self-description are his attempts to *avoid* the wisdom-loving life by hiding from Socrates. I want to turn to another passage in which Alcibiades' attempts at forward progress toward his aspirational target are more evident.

Alcibiades tells how he hungered for Socrates' wisdom and hatched the following plan: "What I thought at the time was that what he really wanted was me, and that seemed to me the luckiest coincidence: all I had to do was to let him have his way with me, and he would teach me everything he knew—believe me, I had a lot of confidence in my looks." Alcibiades goes on to couch his offer of sex-for-knowledge to Socrates in the language of his newfound Socratism: "Nothing is more important to me than becoming the best man I can be, and no one can help me more than you to reach that aim. With a man like you, in fact, I'd be much more ashamed of what wise people would say if I did not take you as my lover, than I would of what all the others, in their foolishness, would say if I did." (218d) Socrates responds, to Alcibiades' astonishment, by complaining that if what Alcibiades is saying is true, he (Socrates) would be getting the short end of the stick:

"Dear Alcibiades, if you are right in what you say about me, you are already more accomplished than you think. If I really have in me the power to make you a better man, then you can see in me a beauty that is really beyond description and makes your own remarkable good looks pale in comparison. But, then, is this a fair exchange that you propose? You seem to me to want more than your proper share: you offer me the merest appearance of beauty, and in return you want the thing itself, 'gold in exchange for bronze.' (218e-219a)

Socrates doubts that Alcibiades really can see what he takes himself to be able to see. If Alcibiades were in a position to fully appreciate the wisdom he claims to have recognized in Socrates, Alcibiades would see that no one who had it would trade it for sex. Someone who sees the beauty of the body as being on par with the beauty of wisdom hasn't apprehended how beautiful wisdom is. When Socrates tells Alcibiades that if he, Alcibiades, were even able to see the beauty of wisdom, he would be "more accomplished than you think," Socrates' point is to show Alcibiades that he is in fact less accomplished than he takes himself to be. Alcibiades'

Alcibiadeanism—with its concomitant arrogance (“I had a lot of confidence in my looks”)—stands in the way of his being able to get his new values into view.

Alcibiades may experience moment to moment variations in the strength of his two points of view, but they are both with him throughout his time as an aspirant, each precluding a full experience of the other. He seems to inhabit both the Socratic value perspective he aspires to acquire, and the honor-loving values represented by the old ways he aspires to escape. He is not merely pretending to care about wisdom for Socrates’ benefit, and then cheerfully pursuing honor on his own time. Nor does he flip-flop between two perspectives, comfortably inhabiting each for however short a duration; he’s not like someone who is drunk and then sobers up, then gets drunk again. Someone who repudiated his past Socratism would not be so haunted by Socratic thoughts. And someone who was fully under Socrates’ spell would not have so much trouble shedding his honor-loving skin.

Alcibiades is, for reasons I will explore below, an especially conflicted aspirant. Aspiration will be smoother sailing for someone who does not make such a sustained effort to escape his own aspiration. Nonetheless, the quality, if not the quantity, of Alcibiades’ conflict is typically aspirational. The work of the aspirant is often marked by some resistance to doing that work. We are much less likely to find such resistance among those engaged in work not geared toward value-change. Someone who is, for instance, building a birdhouse, is unlikely to take apart what he has built, throw away his tools, run away from his project, and then return to working. That’s because his values are fixed throughout the process. Aspirational work calls for courage, since one may feel that one is losing one’s old value system and replacing it with nothing. If we are to get aspirational work into view, we must be prepared to encounter its characteristically tortured and disoriented presentation. At the same time, what it is a tortured presentation of is something that bears a crucial resemblance to building a birdhouse. Both aspiring and building are forms of work, activities to be engaged in and not (merely) experiences to be undergone.

It is not difficult to allow that someone may move from being one kind of person to another. We all drift and change in response to our environment, cultural pressures, etc. Sometimes, we change without even noticing it. The aspirant’s movement is not of this kind, since she actively moves herself. And that means, as I noted above, that she has some *reason* for doing so: aspiration is a *rational* process. Can what Alcibiades is going through in this period of his life be described as a rational process?

Much of this book will be dedicated to articulating both why it is so difficult, and why it is so important, to get the rationality of aspiration into view. Roughly speaking, the problem is that being practically rational involves acting for the sake of some envisioned end—and, as Socrates points out to him, an aspirant like Alcibiades doesn’t fully have in view the end for the sake of which he is acting. Alcibiades cannot fully grasp the value of wisdom, and he therefore cannot fully grasp why it is to be pursued. He offers up his body in exchange for he knows not what, since he doesn’t really see the beauty he thinks he sees. Having an end in view is, of course, merely a prerequisite for the characteristic activities that we usually think of as the mark of the practically rational agent: she reflects on whether this end is, in fact, likely to be achieved by her; she compares it to other ends and decides that it is to be pursued over them; she figures out the best means to achieve it.

The aspirant has trouble engaging in a full way in any of these activities, and that is because she fails the basic prerequisite of acting for the sake of some envisioned end. Her thought about what she is doing cannot be completely clear, both for the negative reason that she has insufficient contact with the value to understand how or why it is to be achieved, and because her mind is positively clouded by the presence of distorting values, such as Alcibiades' love of honor. Alcibiades is not in a position to make value comparisons, such as the comparison between honor and wisdom. He cannot seem to get them both into view at once, as someone might who can 'step back' from two things that he values and decide between them from a position of impartial arbitration. His manner of desiring both honor and wisdom at once is to have one of those points of view dominate, at any given time, and have the other present itself in the form of a recalcitrant impulse or sense of shame or incomplete grasp of the dominant value.

His experience of concomitantly loving honor and wisdom does not come in a form where he might occupy the reflective position of impartial judge between the two values. He does not seem to be trying to 'decide' whether it is better to be honor-loving or wisdom-loving. Rather, he seems convinced that it is better to be wisdom-loving, but unable to feel the force of this conviction in the absence of Socrates. If we look at Alcibiades, we see that he doesn't resemble someone trying to make a decision. Though he clearly struggles, his struggle is not that of someone who has difficulty choosing which of two valuable things he should pursue. The love of honor and love of wisdom that threaten to pull him apart do not present themselves to him in the form of a difficult-to-solve practical problem. He is not trying to figure anything out. He is just trying to become a wisdom-lover, and his love of honor is holding him back. Or, at other times, it seems to be that he is trying *not* to become a wisdom-lover, but finds that honor has lost some of its initial appeal. ("...my heart, or my soul, or whatever you want to call it... has been struck and bitten by philosophy, whose grip on young and eager souls is much more vicious than a viper's and makes them do the most amazing things." 218a) Alcibiades' project is not one of making a decision from the values he already has, it is the project of acquiring the full version of the attenuated values he currently grasps.

I am not going to argue that someone in Alcibiades' position is, despite appearances, in a good position to exhibit the rationality of reflective, reasoned, well-informed decision-making. Instead, I am going to argue that that is only one way in which practical rationality manifests itself.

In order to make a good decision, one must already have the desires (values, preferences, etc.) that supply one with (or just *are*) a conception of the end(s) in the light of which one decides. But how does one come by these very desires, values, preferences etc.? Here are two familiar kinds of answer to that question: (i) Non-agential transformation: I acquire it through external influence/ accident/ genetic predisposition (ii) Self-Cultivation: I acquire it by reasoning that I would better satisfy the desires I have now if I acquired such a desire. Alcibiades' attempt to acquire a desire for wisdom has elements of (i) and (ii), but neither schema fits him perfectly.

A skeptic of aspiration might want to read Alcibiades' story as passive, non-agential transformation along the lines of (i). She would be right to cite the effect of Socrates on Alcibiades' desires and values. Things feel different when Socrates is refuting you; he has a

real influence on the people around him. Such a skeptic might also note Alcibiades' innate talents. Socrates has focused his own efforts on Alcibiades because Socrates, like all of Athens, sees great potential the young man⁸. We should acknowledge that Socrates' influence and Alcibiades' innate talent are relevant to Alcibiades' status as an aspirant—but only as backdrop. They might explain why it is easier for Alcibiades than others to get the value of wisdom into view, but the this skeptic leaves out of the account Alcibiades' activity of (trying to) bring it into view.

Another kind of skeptic reads Alcibiades' pursuit of philosophy as satisfying a set of desires he already has about the kind of person he wants to be. She will point to Alcibiades' belief that he would be better off desiring wisdom, and try to explain his behavior as rationalizable from his antecedent condition. But Alcibiades' belief does not stem from a judgment that such a desire coheres better with, or helps him satisfy, desires he has anyway. His whole problem is that he thinks his deepest, most fundamental desires, values and concerns are misdirected. Alcibiades' project of wisdom-acquisition is both less self-directed and more self-transformative than the process of self-cultivation described by the skeptic⁹. Someone who tries to make her desires more coherent derives guidance from the self she already has; the aspirant must take her bearings from the self she doesn't have yet. More specifically, from some value grasped by that self. Alcibiades' goals of self-improvement require that he fix his attention on the value of wisdom, on seeing *its* beauty. He is surprised to find that his attention is on Socrates, when *he* should be the beloved. The aspiring parent is thinking about her (potential) children, the aspiring music-lover is thinking about the value of music, the aspiring doctor is thinking about the practice of medicine, and her present or future patients, etc. These people are trying to acquire a desire not because something about them demands that they acquire it, but because they see (that they don't full see) that there is something of value out there. Let me step back from the details of Alcibiades' story and place it within the broader philosophical framework of the book.

There is a distinctive form of rationality that pertains to the genesis of desire, a way of reasoning toward desire rather than from it. When we reason from desire, our rationality consists for the most part in making good choices, knowing what is worth sacrificing, taking prudent risks. The excellent decision-maker knows when to jump on an opportunity and when to bide her time. Such rational activity doesn't, of course, entail the presence of a process of *reasoning*. We often make good decisions without doing much thinking; indeed, we sometimes make better decisions by doing less thinking¹⁰. The rational decider is good at making the choices that maximize the satisfaction of her desires (the realization of her values, etc.). In hard cases, she deliberates well; in easy cases, she does not deliberate at all, but immediately decides rationally. Decision theory is that branch of philosophy that offers a formal account of the optimal decision procedure that would be employed by such an agent, be it in the form of consciously articulated reasoning, reflexive rational response, or something that partakes of both. The decision theorist tells us what a rational choice consists in.

⁸ See *Alcibiades I*, esp. 104e-106a

⁹ For further discussion of self-cultivation, see ch. 1, part 3 and ch. 5, part II.

¹⁰ See Holton; and for further discussion of the general point that reflective deliberation is not a cure-all see ch. 3, III below.

Edna Ullmann-Margalit and Laurie Paul have discussed the topic of large, life-altering choices from the point of view of decision theory. Since decision theory presupposes a subject who enters the decision-scenario with (some) fixed core preferences, they argue¹¹ that decision theory cannot tell us how to be rational when the decision's primary import is a fundamental change in those (core) preferences. I find their arguments to be, by and large, sound. I will go through them in the next chapter, though I will draw a very different lesson from their conclusions than they do. Instead of concluding that, because the decision theorist cannot tell us what it would be for an agent like Alcibiades to be rational, there is no fact of the matter about whether he acts rationally, I will argue that agents like Alcibiades—aspirants—exhibit a distinctive form of rationality that is not a matter of decision at all.

Theorists of practical rationality have tended to focus on the question of how we rationally manage the values and desires agents already have. They attend to the case of value-acquisition, to the extent that they do, as a special case of desire-management. There is, of course, a form of desire-acquisition—a secondary form—that is simply a special case of desire-management. My topic, however, is *primary* desire-acquisition. In such a case, the desires one already has are themselves only fully intelligible in the light of the desires one will have. One's present condition is, as it were, a simulacrum of the value condition one hopes someday to be in. If there is such a thing as rational change in one's core values, such a change is not made by trying to best satisfy the values that one already has.

Aspirants often depend on a teacher or mentor figure, but it is a peculiarity of Alcibiades' case how dependent he is on Socrates. In the absence of Socrates, Alcibiades actively opposes the aspirations he feels in Socrates' presence, trying to free himself from the demand to be better. Alcibiades is, on all accounts, a failure at the project of aspiration. This is not the same thing as saying that he does not aspire at all, but it does entail that Alcibiades' behavior, considered as a whole, is irrational. This is, I believe, the point of Socrates' cryptic reply to Alcibiades' speech:

“Alcibiades' frankness provoked a lot of laughter, especially since it was obvious that he was still in love with Socrates, who immediately said to him: “You're perfectly sober after all, Alcibiades. Otherwise you could never have concealed your motive so gracefully: how casually you let it drop, almost like an afterthought, at the very end of your speech! As if the real point of all this has not been simply to make trouble between Agathon and me! You think that I should be in love with you and no one else, while you, and no one else, should be in love with Agathon—well, we were *not* deceived; we've seen through your little satyr play.” (222cd)

What might strike the reader as an unfeeling or even clueless response to Alcibiades' impassioned confession is, I believe, meant to function as a rebuke. Socrates charges Alcibiades with the ulterior motive of “getting between” himself and Agathon. It is striking, however, that Socrates does not suspect Alcibiades of seeking to court Agathon's affections. Socrates thinks that Alcibiades is trying to be the *lover* of Agathon, and the *beloved* of Socrates. The narrator remarks, of Alcibiades, that “it was obvious that he was still in love with

¹¹ Here I am glossing over a distinction between the two views. It is Ullmann-Margalit who has the view described here, while Paul hopes to rescue a decision theoretic approach by appealing to second order preferences. I argue that her rescue does not work, and so her critique leaves us in the same position as Ullmann-Margalit's.

Socrates,” but Socrates notes that Alcibiades is fighting this condition. Alcibiades wants *to be in love* with someone else, to pursue the up-and-comer Agathon, whom he came to the party to crown with laurel wreaths (“I want this crown to come directly from my head to the head that belongs, I don’t mind saying, to the cleverest and best looking man in town.” 212e). Alcibiades also wants to be loved by Socrates, for that would mean that Socrates would take his ‘proper’ place: older, uglier lover (*erastēs*) to a young, beautiful boy (*erōmenos*)¹². He wants Socrates to want him for his young, beautiful body, so that he can stop loving Socrates and himself pursue the even younger and better looking Agathon for *his* beautiful body.

Socrates accuses Alcibiades of trying not to love Socrates. ‘Loving Socrates’ has already emerged, through Alcibiades’ speech, as a matter of striving for virtue and wisdom. Socrates’ response to Alcibiades sexual overtures was to invite him as a partner in the shared project of acquiring the wisdom he thinks they both lack: “In the future, let’s consider things together. We’ll always do what seems the best to the two of us.” (219b) Likewise, Socrates’ conviction that Alcibiades is evading or resisting the work of love amounts to an accusation that he is unwilling to seek the value he aspires to have. Alcibiades is aspiring irrationally, because he is trying to be better, and trying not to be better, at the same time. He interrupts his own pursuit in such a way as to make it discontinuous, incoherent, and ultimately fruitless.

The reader might be wondering why I would choose, as a model of aspiration, someone who is not a model aspirant. The answer is that by modeling *irrational* aspiration Alcibiades reveals to us that the distinction between rationality and irrationality does indeed have application, even in such murky waters. In hiding from Socrates (“My whole life has become one constant effort to escape from him and keep away,” 216b) Alcibiades is also hiding from the part of himself that hungers for something more than honors and power. And this is surely part of what Plato is trying to expose by having him tell his story.

During the drinking party, Alcibiades is still Athens’ golden boy, but by the time the dialogue was written, Athens had witnessed Alcibiades’ spectacular downfall as well as the trial and death of Socrates. The disastrous Sicilian expedition urged by Alcibiades in the year after the symposium was the beginning of the end, leading to Alcibiades’ betrayal of Athens to the Spartans, whom he in turn betrayed to the Persians, before briefly returning back to Athens until, having been exiled once more, he was executed by a party sent by the Athenian leadership. The 19th century historian and lexicographer William Smith describes his end aptly: “Thus perished miserably, in the vigour of his age, one of the most remarkable, but not one of the greatest, characters in Grecian history. With qualities which, properly applied, might have rendered him the greatest benefactor of Athens, he contrived to attain the inhumane distinction of being that citizen who had inflicted upon her the most signal amount of damage.” (p.376)

Give its dramatic date, and the date when it was written, Plato’s depiction of Alcibiades’ speech must be read in the light of his trajectory¹³. Arguably, the core of the animus fueling Socrates’

¹² Socrates’ comment thus picks up on one of Alcibiades’ remarks toward the end of his speech: “He has deceived us all: he presents himself as your lover, and, before you know it, you’re in love with him yourself!” (222b)

¹³ A reading encouraged by the fact that Plato inserts the tale of the dinner party into a much later narrative frame. He has Apollodorus recount the events of the party to an unnamed friend some years after the deaths of both Socrates and Alcibiades.

accusers is the negative influence they judged him to have on the noble youths who flocked to his company¹⁴. We do not know what was said in the speech that preceded the one Plato presents in the *Apology*, but the case against Socrates certainly could have referenced Alcibiades as exhibit A. In the *Symposium*, Plato may be offering us an alternative moral to the story of the fate of Alcibiades. Plato seems to be tracing the disastrous outcome of Alcibiades' life to the fact that he didn't try hard enough. He had available to him the raw materials (talent, influence, the help of Socrates) for a meaningful life, and let himself get sucked back into a life he could see was empty. Alcibiades' practical irrationality consists in his culpable failure to aspire sufficiently.

In the final section of this book, I will argue that the theorist of aspiration is well-placed to vindicate moral responsibility for one's self by solving a puzzle about how we can be 'causes of ourselves.' The aspirant herself, and not (only) her parents, teachers, mentors, culture etc., is morally responsible for developing into the kind of person she ends up becoming. The reason Socrates didn't "make" Alcibiades virtuous is because that is not something one person can do to another. The work of appreciating the value of, say, wisdom, is not work that someone can do for or to you. As Socrates was constantly telling people, virtue is not teachable.

IV. Self and Value

There is a conceptual connection I have been taking for granted throughout this introduction. I have described aspiration as, on the one hand, the process by which we acquire values, and, on the other hand, the process by which we become a certain kind of person. As I will continue to identify these processes throughout the book, a word of explanation is in order. First, I should note that I use words such as "self," "identity," "character" and phrases such as "who a person is" and "becoming someone" and "the kind of person one is" in a strictly ethical as opposed to metaphysical way. I am, thus, setting aside metaphysical questions as to what, if anything, allows a person to remain one single object over the course of the many changes she undergoes from birth to death. I am not discussing a metaphysical question about personal identity over time, but rather an ethical question about a person's true or real or deep self. This self is composed of those features of a person that have ethical significance—they are the features in virtue of which you are praise- or blameworthy, beloved or hated. When you are proud or ashamed of some feature of yourself, that feature has some ethical significance for you. Which features are those? They will differ from person to person, and they depend at least in part on what the person *takes* to have ethical significance.

Consider the following list of facts about a person: health, heritage, race, religion, gender, hobbies, profession, fashion sense, physical appearance, citizenship, how her home is decorated, her sexual orientation. Most of us will take some, but not all, items on this list to reflect "who I really am." One atheist may be passionately committed to her atheism, devoting herself to the cause of undermining religious authority; another rarely gives the matter a second thought. One

¹⁴ This charge appears both on the lists of official charges brought by Anytus, Meletus and Lycon against Socrates (*Apology* 24bc) and on the charges Socrates attributes to his "old accusers" (*Apology* 19c). Additionally, the question of the education or corruption of the youth is the bone of contention between Socrates and Anytus in the *Meno* (89e and following), and between Socrates and Meletus in the *Apology* (24d and following).

person treasures her knowledge of multiple languages, taking care to, e.g. pass them down to her children, whereas another is indifferent to the fact that he possesses this capacity. The fact that so-and-so is your next door neighbor might be a really important fact about your life—you'd be devastated if she moved—or you might barely know her.

Of course an agent's indifference to a fact is compatible with its having profound ethical significance for her identity. Someone who is indifferent to the needs of her immediate family members or friends, or to the dignity and equal worth of other human beings, manifests an ethically significant form of indifference. Nonetheless, even in these cases the ethical significance is a fact about what she takes to be important: a fact about, as it were, the negative space around her caring. Likewise, if you judge that someone cares too much about, e.g., what others think of her or how she looks, you might fault her for the shape of her concerns. All of these facts—both the facts that have ethical significance (in part) because the person endows them with such, and the facts that have ethical significance because they reflect failures in her endowment-system—are facts about a person's *values*.

When I speak of a person's ethical self, I am really talking about her values. When I speak of a person making profound changes to herself, I am talking about her changing her values; and when I speak of the aspirant's awareness of a deficiency in herself, I am describing a deficiency in respect of value. But is it really true that a person's values are the ethically deepest and most important facts about her? Giving a complete defense of these theses about ethical identity is beyond the scope of this book, but I do want to say a few more things to defend the intuitive plausibility of the ethical association between a person and her values.

This association is reflected in our ordinary, everyday speech and thought about ourselves. When we say "I'm not that person anymore" we usually mean that the values we have now differ from those we had then. When we promise a loved one to whom we have caused profound hurt that we will change, we mean that we will change not only our behavior but the values from which that behavior springs. We feel shame at being discovered to be in some condition when being in that condition amounts to a failure to live up to our values; and, correspondingly, when we feel pride it is because we recognize the fact that we have lived up to our values. When we contrast superficial features of a person, such as his appearance or social status, with "what really matters about a person," we are referring to a set of features of the person that either are, or at least include, the person's values. So, for instance, when Dr. King enjoined us to judge people "not by the color of their skin, but by the content of their character," he was contrasting a superficial feature of a person with a deep one. It is a mistake to judge a person by his skin color, but it might not be a mistake to judge him by the fact that he, for instance, unduly esteems his own skin color, and takes it as grounds for treating those with skin of a different color as his moral inferiors. The latter is a fact about a person's (defective) values, and those are the things that comprise a person's character.

But *why* is it the case that a person's values are so central to who she is? To answer this question, we must first explain what values are. Is valuing a form of desiring, as David Lewis (1989) and Gilbert Harman (2000) have argued? Or is it, as Michael Smith (1992) maintains, a kind of belief? Samuel Scheffler (2010) has argued against such reductive accounts of value, and in favor of an account of value on which it involves "a complex syndrome of interrelated

dispositions and attitudes, including (at least) certain characteristic types of belief, dispositions to treat certain kinds of considerations as reasons for action, and susceptibility to a wide range of emotions.” (Scheffler p.4) Jay Wallace (2013) and Niko Kolodny (2003) have also espoused a theory of valuing on which it constitutes a response to a variety of different kinds of reasons: the valuer sees reasons to believe that what she values is good (or valuable or worthy), as well as reasons to do things in relation to it, and to have certain feelings at the prospect of engaging with the value, losing the value, etc. I follow Scheffler, Wallace and Kolodny in accepting such a “hybrid account” of value, which I discuss in chapter III. On such an account, when we identify someone’s values we identify objects around which her ethical self is organized: the valued object elicits from the agent a response that is the product of the cooperation of the various cognitive, motivational and affective elements of her agency. Values are, therefore, a nexus of the person’s various agential functions; when we know a person’s values, we know what objects out in the world she is, as we say, “all about.”

Finally, I want to address a worry about associating aspiration with the self. Someone who accepts the association I have just described, between aspiration and value-acquisition, might nonetheless object to the idea that what aspirants are engaged in a kind of self-making, self-shaping or self-creation. For she might think that aspiration ought to be directed, first and foremost, not at the fact that I will have a certain value but rather at the valuable object itself. Let me return to the case of the aspiring doctor. The objector I am imagining would insist that the aspiring doctor is directed at helping people with their medical problems rather than at discovering the value of helping people with their medical problems. Even granting that the medical student is open to such a learning experience, the objector I am imagining maintains that what she is really ‘all about’ is helping people, not changing herself. The new set of preferences she acquires at the end of her training would then not be her aim, but a consequence of what she aspires to.

Consider, however, the importance of hands-on experience in coming to comprehend the many spheres in which the agency of doctors is operative: there is preventative care, bedside manner, surgical precision, enabling people to make informed choices, consulting with colleagues over a diagnosis, making lifelong connections with patients and their families, etc. The aspiring doctor, like those of us who are not doctors, has only a schematic understanding of the helping activities she will perform in each of these spheres. For instance, she may not have thought much about how she will navigate the dilemma of supporting a patient’s autonomy while at the same time advising them in making a medically sound decision. For the main way in which one would come to understand that dilemma, and the distinctive kind of medical help one provides in offering advice, is by working with actual patients to guide them through the decisions they must make.

The medical student cannot simply aim to realize the value of helping people, or some more specific value such as that of helping people make good medical decisions, because she does not have a firm grip on what she would be realizing. The experienced doctor, by contrast, aware that she is entering a consulting room with a patient who has a difficult choice before him, can straightforwardly possess the relevant aim. She thinks to herself, “I want to help this patient make a good decision without telling him what to do.” I do not want to deny that the aspiring doctor aims at the goal of helping people. Rather, I am claiming that aiming at this goal when

one's knowledge of it is limited just is a matter of trying to learn what that goal amounts to. This is not because the agent in question is self- as opposed to world- directed, but because the only shape that her contact with the value can take is an educational one. She comes into contact with, or aims at, the value not by realizing it but by learning it. In such a case one's value-directed activity is simply identical with a self-directed activity. This learning constitutes a change in who she is, which is to say, a change in what she values.

* Chapter 1: Decision Theory and Transformative Choice

Decision theory is a branch of study lying at the intersection of philosophy and economics, devoted to the analysis of rational preference structures and rational choices. It is the job of the decision theorist¹⁵ to articulate a procedure for deliberating well between any set of options. Decision theorists explain how one makes the right decision. But are there choice situations for which no such procedure can, in principle, be supplied?

Edna Ullmann-Margalit's 2006 paper "Big Decisions: Opting, Converting, Drifting" and Laurie Paul's 2014 book *Transformative Experience* take up the question of whether decision theory can offer us an account of the rational navigation of life's major crossroads. In addition to familiar examples such as attending college, having children, getting married or choose a career, both authors offer us a rich medley of more idiosyncratic examples of life-altering decisions. Ullmann-Margalit considers the early Zionists who had to leave their eastern European heritage behind in order to "become the new Jews of their ideals." (p.160) Paul discusses the difficult choice a member of the Deaf¹⁶ community faces when offered the surgical procedure that would (partially) cure her infant child's deafness, but also thereby alienate him from the Deaf community. Their question, about all such agents, is what it would be to count any of them as having been made rationally.

Ullmann-Margalit distinguishes such 'big decisions' (as she calls them) from both the medium-sized decisions that she takes decision theory to be well-placed to handle, and 'small decisions' such as choosing one token (e.g. a box of cereal) over another nearby token of the same type. In a small decision, an agent finds himself choosing—or as Ullmann-Margalit (1977) calls it in an earlier paper, "picking"¹⁷—between near-identical items. Why should I choose this cereal box instead of the one next to it? Perhaps there is simply no answer to this question: Ullmann-Margalit argues that the distinction between rational and irrational choice does not get a grip in small decisions, because they are "selection situations without preference."¹⁸ We will, perhaps, not be too troubled to hear that decision theorists offer us no guidance in making small decisions, given that people making those decisions typically are not looking for guidance. We don't, in ordinary circumstances, agonize over the question of which cereal box to select. The silence of decision theory in respect of big decisions would, however, be much more troubling. We *do* agonize over the choice to have children, or over major career decisions, and other such forks in the road of life. Ullmann-Margalit argues that it is nonetheless no more possible to draw the distinction between deciding rationally and deciding irrationally when one is deciding between careers, than when one is deciding between cereal boxes. Decision theory can only

¹⁵ Though not necessarily the job of *every* decision theorist. First, because, as Ullmann-Margalit points out, some decision theorists set out to describe how agents actually decide rather than how they ought to decide; second, as Pettigrew points out in his comments on Paul's book, some of those in the latter group are concerned only with what it takes for a set of preferences to be well-ordered rather than with articulating a decision procedure that agents might incorporate into their deliberations.

¹⁶ Like Paul, I follow the convention of reserving "deaf" with lowercase "d" for the biological property of being unable to hear, and write "Deaf" with an uppercase letter to describe the culture and community that exists among many deaf people.

¹⁷ "When preferences are completely symmetrical, where one is indifferent with regard to the alternatives, we shall refer to the act of taking (doing) one of them as an act of *picking*." (1977, p.757)

¹⁸ Ullmann-Margalit and Morgenbesser 1977 p.758; in her later paper, she calls them "decisions without preferences" (Ullmann-Margalit 2006, p.171)

help us answer the ‘medium-sized’ questions such as whether or not we should buy a car¹⁹.

The problem discussed by Paul and Ullmann-Margalit can be stated at a very high level of generality, one not tied to any particular conception of decision theory. The decision theorist asks an agent placed before any choice to consider each outcome’s likelihood of occurring, and her preferences with respect to those outcomes. In order to make a rational decision to ϕ (for the sake of A) instead of ψ -ing (for the sake of B), one needs to know whether (and by how much) one prefers A to B, and whether (and by how much) it is more likely that ϕ -ing produces A than that ψ -ing produces B. The problem is that because one (or both) of the options promises a substantial change in preferences, the agent doesn’t have a single, stable set of preferences which could provide the input for the decision procedure.²⁰ Over the course of a period of time during which my desires are unstable, there may be no fact of the matter about which option would better fulfill them. Decision theory assumes relatively²¹ static preferences, and an agent making a big choice confronts a radical shift in her preferences.

Paul and Ullmann-Margalit make this point in slightly different ways. Ullmann-Margalit examines the rationality of a big decision from the outside, posing the question as to what standard we should use in evaluating the rationality of someone else’s big decision. Paul is interested in inspecting a big decision from the point of view of the agent making it. She asks us to consider what it would take to count ourselves justified in our own choice. I think we can gain insight both from Ullmann-Margalit’s third-personal perspective on the problem, and from Paul’s first-personal one. I begin with Ullmann-Margalit.

I. Ullmann-Margalit on Old Person vs. New Person

Ullmann-Margalit calls making a big decision “opting,” and describes the problem thus:

New Person is now, by hypothesis, a transformed person. Opting transforms the sets of one’s core beliefs and desires. A significant personality shift takes place in our opter, a shift that alters his cognitive as well as evaluative systems. New Person’s new sets of beliefs and desires may well be internally consistent but the point about the

¹⁹ In her earlier paper, she rightly points out that the distinction between small and medium sized decisions is a matter of attitude (p.780): a child might take the decision between (what look to the adults to be) two identical pieces of candy to be a medium sized decision; whereas an unusually indifferent adult—perhaps she is very wealthy, perhaps she is depressed—might see the purchase of a car or house as a small decision. We can elide this point by simply assuming that we are describing ordinary adults having the reaction we might expect to some choice situation. Thus when I describe a cereal box choices as small, a car-purchases as medium sized, and choices concerning marriage, pregnancy, and career as large, I presuppose that those choices are viewed from the perspective of an ordinary adult. I acknowledge that the situations from which I, like Paul and Ullmann-Margalit, draw my examples of ‘big decisions’—marriage, emigration, pregnancy, career—can present as small or medium sized choices for extraordinary people, or for ordinary people in extraordinary circumstances.

²⁰ Here I subsume questions of the agent’s tolerance of risk under the aegis of preference; I also set aside the question of whether there are other decisions that agents cannot navigate because they are ignorant as to the relevant probabilities. Paul argues that the value-ignorance of agents making large choices cannot be handled by existent methods of accommodating ignorance about probabilities. (p.29ff.)—though this argument rests on the assumption, challenged below, that such agents have not already begun the transformative pursuit in question.

²¹ The decision theorist can manage small shifts in preference that occur against a stable background of core preferences (see Elster); hence Ullmann-Margalit’s emphasis on the fact that a decision is only ‘big’ in the relevant sense if it is ‘core-affecting.’

transformation is that inconsistency now exists between New Person's system of beliefs and desires, taken as a whole, and Old Person's system taken as a whole. I am not questioning his ability to actually make a choice, or his ability subsequently to assess himself as happy (or unhappy) with his choice. The question I am raising is whether it is possible to assess the rationality of his choice, given that this choice straddles two discontinuous personalities with two different rationality bases." (p.167)

Here is one of the examples with which she fills out the schema of opting:

"I was told of a person who hesitated to have children because he did not want to become the 'boring type' that all his friends became after they had children. Finally, he did decide to have a child and, with time, he did adopt the boring characteristics of his parent friends—but he was happy! I suppose second order preferences are crucial to the way we are to make sense of this story. As Old Person, he did not approve of the person he knew he would become if he has children: his preferences were to not have New Person's preferences. As New Person, however, not only did he acquire the predicted new set of preferences, he also seems to have approved of himself having them. How are we to assess the question whether he opted 'right'?" (fn. 10)

Ullmann-Margalit thinks that we cannot assess the rationality of this choice, because in order to do so we would have to make an arbitrary selection from the two points of view from which we might assess it. If we adopt the value-perspective of Old Person, the choice to have children was the wrong one: it frustrates Old Person's desire to avoid being boring. On the other hand, if we adopt New Person's point of view, the choice was rational: it satisfies New Person's core preference for conventional family life. There is no neutral perspective from which we can answer the question as to whether his choice left him 'better off' than he was. Ullmann-Margalit takes the question with which this example ends to be rhetorical: decision theory offers us no way to answer them, and they are therefore unanswerable. She concludes that the 'opting' we do in big decisions suffers from the same rational unassessability as the 'picking' we do in small ones. She takes it that we are no more able to evaluate the rationality of Old Person's choice to have children than his choice of this box of cereal over that one. Both big decisions and small decisions are cases where "reasons cannot prevail." (p.171)

She does not charge someone in this case with irrationality, but rather arationality: "In order to be irrational about something there must also be a rational way of going about it, and the rational way of going about opting is what I am here questioning. (p.168)"

II. Paul on Deciding to Become a Vampire

Paul's book explores the same themes as Ullmann-Margalit's essay, but from a first personal perspective. Most of her book is dedicated to articulating the difficulty of taking account of the value of the prospective experiences that would be the product of what she calls our "epistemic and personal transformations." (p.17) Her guiding example is one in which the reader is offered the chance to become a vampire: "As a member of the undead, your life will be completely different. You'll experience a range of intense, revelatory new sense experiences, you'll gain immortal strength, speed and power, and you'll look fantastic in everything you wear. You'll also need to drink blood and avoid sunlight." (p.1) Paul's claim is that because you cannot

know what it is like to be a vampire until after you are bitten, you cannot make an informed choice in the traditional way. In particular, nothing you (think you) know about vampiric life is relevant to your choice.

Your ignorance is unchanged by the fact that, as Paul imagines the scenario, others have walked the road before you: “all of your friends, people whose interests, views and lives were similar to yours, have already decided to become vampires. And all of them tell you that they love it.”

(p.1) You might reasonably expect that you’ll feel happy with the choice if you make it; nonetheless, you shouldn’t think you have any idea of what you’re getting yourself into. She considers whether someone could approach such a choice by conducting a scientific study of happiness among those who have, and those who have not, made the choice in question. Such studies have in fact been done on the childbearing decision, and Paul cautiously concludes that the research points in the direction of the conclusion that childlessness is the rational choice:

“...if you are prepared to ignore all of your subjective assessments about what it would be like for you to have a child, and choose solely on the basis of the empirical research, if you want to maximize your expected subjective value, the research (to the extent that there are clear results) suggests you should not have a child. If we simply follow the dominant empirical conclusions of the experts, it seems that anyone who wants to make a rational decision about parenthood that is based on maximizing expected subjective value must either suspend judgment (given the lack of a clear consensus on the results), or should actively choose to remain childless.” (p.87)

Ullmann-Margalit would object that studies of this kind must adopt either the one or the other point of view on happiness; and that as long as the choice is sufficiently transformative there is no common denominator that could encompass both perspectives on value. For instance, the person who embraces the perspective of a parent will argue that the scientist’s conception of happiness has no way of taking into account the distinctive pleasure one takes in one’s children.²² Ullmann-Margalit’s thought is that the relevant options cannot be impartially compared²³.

Paul also objects to basing one’s decision on such studies, which she calls “impersonal, “big data” reasoning” (p.88), but on different grounds. Her worry is not that such reasoning must always prejudice the answer in favor of one of the two perspectives, but rather that one sacrifices autonomy by approaching a big decision choices in such a detached spirit:

“Unless robots have taken over the world when humans weren't looking, for many of us, this is

²² For a particularly eloquent instance of such a defense see Jennifer Lawler’s essay on parenting her disabled daughter: “If someone had told you ahead of time what was going to happen now? Baby, you would have been on the next plane to Bolivia and fighting extradition every step of the way. But they didn’t tell you ahead of time, and by the time you figured out that being her mother was going to make your life look like a nuclear bomb had detonated in the middle of it, it was too late, because she’s your daughter and you loved her even before she was born, so you’re a little biased and you can’t always see her clearly, and what you see is a high-spirited, ebullient girl with a stubborn streak, and other people see a slow-moving, cognitively-impaired kid who can’t be budged once she makes up her mind.... And you hug her hard, but she’s used to that, too, and she lets you, and even lets you sing along without complaining (“this time only, mom!”), and you are lucky, probably the luckiest woman living, and happier than you have ever been, but not in any way an academic would understand, or even conceive. Your joy is bigger than the universe and contains all the sorrow of a lifetime, and has nothing whatsoever to do with feeling sufficiently rewarded for your work.” cf. Harman 2014, fn. 5.

²³ The line of thought I’m attributing to Ullmann-Margalit here is one that I develop in chapter three below. There I describe conflicting evaluative perspectives that do not admit of a deliberative resolution.

an untenable way to approach choices involving our personal goals, hopes, projects, and dreams. In other words, in today's society, when making important personal choices, we want to consult our own, personal preferences and to reflect on what we want our future lives to be like as part of assigning values to outcomes. It is simply unacceptable to be expected to give up this sort of personal autonomy in order to make decisions about how one wants to live one's life."

Her contention is that we feel we don't make our choices rightly unless we make them from a personal point of view, by asking what this experience will mean *for me*. But in cases of big decisions, we cannot answer this question. All we know is that the decision will produce a new and revelatory experience.

III. Self-cultivation vs. Transformation

It is important to note that only some examples of emigrating, motherhood or becoming a vampire will serve Paul and Ullmann-Margalit's purposes. We could imagine a dutiful daughter who moves to Israel in order to fulfill a promise to her mother; a younger sister who, having observed an older sister's pregnancy and childraising firsthand, feels she has an excellent grip on what those changes will mean for her; a fan of vampire movies who (believes she) wants nothing more than to live an emotionally vacuous, immortal, and fashionable life. The dutiful daughter recognizes that her life may become more difficult as a result of her choice, but counts this as a side-effect of a decision grounded, fundamentally, in duty. The younger sister and the movie lover may be deluded as to the experience of motherhood or vampiricity; nonetheless, the important point is that they took themselves to know what they were getting themselves into. None of these people makes a transformative choice in Paul's sense or a big decision in Ullmann-Margalit's sense. Likewise, the fact that motherhood or emigration ends up making no mark on someone's preferences doesn't prevent her choice from qualifying as big/transformational, so long as she did not know this would be the case. And of course many decisions we make to change the shape of our lives will *not* be large or transformational in Paul and Ullmann-Margalit's sense. For we may feel, rightly or not, that we know well in advance what we are getting out of, e.g., getting a new job.

Among these various kinds of cases that will not serve Paul and Ullmann-Margalit's purposes, it will be useful to have a label for those that are aimed specifically at self-change. I will call it *self-cultivation* if someone decides to, e.g., join a gym in the expectation (and hope) that she will thereby, over time, engender in herself an inclination to exercise. Or rather, I will call such a decision self-cultivation so long as that inclination serves her current stable interest in living a more healthy life. Likewise, someone might take a speed-reading class, change her sleep-habits, or enter psychotherapy, all the while foreseeing, and approving of, the change in preferences that these decisions will occasion. I can now see (or think I can see) that I want to be the kind of person who takes pleasure in exercise, who is motivated to do more reading, who is inclined to wake up early, or who has a less anxiety-grounded preference structure. In changing myself in these ways, I am driven by my current preferences. For I see the having of those new preferences as a way of satisfying current and more basic preferences to, e.g., live a long life, get more work done, etc. Likewise, there are cases in which we cultivate a taste for, e.g. opera or cigarette smoking or gourmet food in order to reap the social rewards of such an interest in one's peer group. These cases do not raise the puzzles described by Paul and Ullmann-

Margalit, because the agent enters the choice fully equipped with the resources to appreciate and value the person she is making herself into. So: what could possibly ground a preference-changing decision if those preferences do not support, but rather threaten, one's basic antecedent preference structure?

IV. Privileging Second-Order Preferences?

At the end of her book, Paul proposes that the very fact that our knowledge that the value-experience in question will be *new* might itself form the basis for a rational transformative choice. She suggests that we can ground these decisions on whether we want things to change or stay the same:

"We must embrace the epistemic fact that, in real-life cases of making major life decisions in transformative contexts, we have very little to go on. To the extent that our choice depends on our subjective preferences, we choose between the alternatives of discovering what it is like to have the new preferences and experiences involved, or keeping the status quo. If we decide to choose this way, when facing big life choices, the main thing we are choosing is whether to discover a new way of living: life as a parent, or life as a hearing person, or life as a neurosurgeon, and so forth; that is, we choose to become the kind of person—without knowing what that will be like—that these experiences will make us into. Or, because we value our current preferences more highly than we value the (mere) discovery of new ones, we reject revelation." (p.122-3)

Paul recognizes that the second-order preference for preference change (or stasis) is one that itself may change as a result of the choice. She contends that it is nonetheless rational for the agent to privilege her current second-order preference and use it to decide what she wants to do²⁴. This procedure, she claims, at least avoids the *irrationality* of "trying to decide based on the character of the particular subjective values of the lived experiences involved," for it allows us to avoid "let[ting] illicit, unjustified assumptions about what it will actually be like to be a vampire....infect our decision procedure." (p.121) Paul's account of rational transformative choice is sketched in only a few pages of her final chapter; it is, presumably, an account whose details she plans to elaborate in future work. I want to offer an objection to the approach she proposes, at least in its current form.

First, let us note that it only applies to a subset of big decisions. Ullmann-Margalit (p.161) draws a distinction between opting (yes, no) and opting (A, B) that I find quite helpful for analyzing Paul's solution. In opting (yes, no) the agent must choose between, for instance, becoming a vampire or not, becoming a parent or not. In opting (A,B) both options are transformative: she is choosing between two careers, or between having a child and travelling the world, or between marrying suitor A and marrying suitor B. Paul's suggestion that the agent choose based on whether or not she wants a change does not offer any guidance to the agent in the opting (A, B) situation. For in those cases, both choices will satisfy (or frustrate) whatever second order preference one has. In a footnote, Paul effectively acknowledges that in such

²⁴ Then why isn't it rational for her to similarly privilege her current first order preferences? There is a worry here akin to that raised by Watson (*Free Agency*, p.217-219) against Frankfurt's higher order desire theory. From the fact that one mental state is of a conceptually higher order than another nothing follows about the relative authority between the two.

cases we may be unable to choose on a rational basis: “perhaps we can only pick a career, not choose a career, in the sense of “picking” developed by Ullmann-Margalit and Morgenbesser (1977).”

Beyond the fact that it is theoretically unsatisfying to solve only part of the problem, there is a question whether the difference between the two kinds of cases is strong enough to support separate treatment. For there is a problem as to whether even the opting (yes, no) cases offer us the prospect of stasis. As Ullmann-Margalit emphasizes, it seems to be a feature of big choices that “the choice not made casts a lingering shadow.” If we choose not to have children, we do not go on exactly as before. For now we are living the life in which we have chosen to live without children. In choosing not to become vampires, we are bringing our own humanity to the forefront of the rest of our lives. To choose but one of the myriad potential practical repercussions of such “stasis”: if one’s friends have all become vampires or engrossed in their new children, the choice to “remain as is” may entail getting new friends.

Even when the ‘no’ option lacks external practical consequences, it typically has internal ones. As Ullmann-Margalit points out: “The rejected option enters in an essential way into the person’s description of his or her life. The shadow presence maintained by the rejected option may constitute a yardstick by which this person evaluates the worth, success or meaning of his or her life.” Opting against transformation can turn a previously unnoticed property of oneself into a focus of one’s identity: one might proudly wear the banner of childfree by choice, mortal by intention. Our subsequent experience of value may well be colored—transformed—by the fact that the other option was available, and rejected. The stasis that Paul’s no-opter seeks may not be available to any subject of a big decision, since it seems that even deciding against changing can be transformative²⁵.

But this is to cast doubt on whether Paul’s procedure is applicable to many actual big choices. We can certainly concoct agents for whom it will be applicable. (Someone who knows she will have the option, after she opts “no,” to take an amnesia pill really can bring it about that life goes on much as it did before.) I want to set aside worries about whether we can apply the second-order preference solution and ask whether we ought to apply it. I think that even if we could make (yes, no) decisions by looking to second order preference, we should not do so.

The fact that Paul’s solution cannot help us opt (A,B) is connected to the fact that it is troublingly one size fits all. If she is right, someone ought to think about whether or not to adopt a child in just the way that she would think about embarking on a career as an ultra marathon runner, getting a cochlear implant (for one’s child), or emigrating from one’s Soviet homeland to build the state of Israel. If one knows that one wants to do one of these things, one knows one wants to do any of them! This follows from the fact that one’s desire to have a revelatory experience cannot (rationally) be conditioned on the way one expects that particular experience to be. In either case, one is setting off on a new adventure that will change one’s preferences in ways one, if one is rational, will recognize that one cannot predict. One should

²⁵ It isn’t even obvious that not having children is *less* transformative than having them. There are people for whom having children would be the conventional and expected choice, and the decision to remain child-free would, in ways they cannot fully anticipate, set them on a radical new path in which they have broken from their communities. More generally, if I cannot know in advance what a transformative experience will be like, I also cannot know how transformative it will be.

make (either) decision if one wants to know what it is like to have that new experience.

Couldn't Paul claim that one wants to discover *what it is like to be a mother* without wanting to discover *what it is like to be an Israeli pioneer*? The problem is that, on Paul's view, the relevant difference between these options is subjective and experiential, and it is precisely these features that Paul cautions us against importing into our decision. Again, it is for this reason that the second-order method does not help us choose between pioneerhood and motherhood—because it must abstract from the details of any particular experience to the question of whether it is new or old. Notice that this problem does not occur in cases of self-cultivation. I can use the fact that I prefer to live longer to ground a second order preference for preferring to exercise, and on this basis decide to join a gym. But in this sort of case is a case of self-cultivation precisely because the second order preference is grounded in a first order preference. I prefer to prefer to exercise because I prefer to be healthy. A free-floating second-order preference, such as the preference for preference-change, does not provide a similarly stable anchor for choice.

I have not ever made a big decision using Paul's method, and I don't know anyone who has. This may be because the people I know haven't had a chance to, since too few big decisions offer a stasis option. Or it might be because we do not take this to be a good way to make the decision. In any case, I would be surprised to hear someone articulate a big decision, such as the decision to have a child, primarily in terms of whether she wants to change or stay the same. We do not seek to transform our preferences for the sake of doing so, but because we want to have the new set of preferences. We certainly know that by choosing to become parents or pioneers we will have new experiences; nonetheless, our reason for opting yes is not the newness but the particular value of those experiences²⁶. This is why a reason to have children is not also a reason to become a vampire, or move to Israel. Ullmann-Margalit and Paul are right that a person cannot get a full grip on the kind of goodness the new experience offers her from her current point of view—but that, I will argue, is exactly why she needs to aspire.

V. Reasoning Toward Desire: Rational Self-Transformation

Paul and Ullmann-Margalit show us how difficult it is to justify a transformative decision from the perspective of the way of thinking and valuing that that decision is going to transform. It is striking that, though both Ullmann-Margalit and Paul cast their claims in relation to decision theory, none of the technical details of such a theory is of relevance to either's argument. Ullmann-Margalit needs only the "core insight" that an action is rational if it is "the best way of satisfying the full set of the person's desires, given his or her set of beliefs formed on the basis of the (optimal amount of) evidence at their disposal." (p.164) Paul works from the fact that "the normative decision maker should choose the act that has the highest expected value." (p.21) In both cases, the point of invoking decision theory is to establish that the desires and value-expectations which serve as the input in the agent's reasoning are the desires and expectations she has *before* making the decision. "Decision theory" functions as a placeholder²⁷

²⁶ Here I'm staying in Paul's metaethical framework, in which all value is translated into the subjective value of experiences. I do not think that this is the right way to think about value, but that issue is immaterial here.

²⁷ Paul concurs, to some extent: "the main problem with truly transformative choice is not a problem in formal epistemology; it is a problem in formal phenomenology." (p.156) She sees the problem as lying in the fact that we cannot know what it is like (to be a parent, or pilgrim, etc.) in advance.

for any mode of reasoning that takes its bearings from the desire or value condition we are currently in.

The project of this book is to expose the possibility of reasoning that engages in the first instance not with what we *do* desire or value at the time of the reasoning, but with what we *will* desire and value. An agent can exhibit rationality in working to apprehend the value of the life she seeks to have. I am going to challenge the prevailing assumption that basic or fundamental preferences (desires, values, etc.) are the kinds of things you can only reason *from*, by exposing a way we have of reasoning *toward* them.

Recall Paul's cautionary advice to the prospective mother, pioneer or vampire against the inclination to "infect" her decision procedure with "illicit, unjustified assumptions about what it will actually be like to be" the person she's considering becoming. (p.121) Paul's point is not that one should incorporate licit, justified beliefs about what it will be like into one's deliberations. Her point is that since we cannot know what it will be like, *all* of our antecedent thought about what it will actually be like will be illicit, unjustified assumptions. Her account of the difficulty of rationally becoming a mother or vampire rests on a claim about the unknowability of a wide range of transformative experiences. She discusses the transformative experience of seeing a color for the first time, of tasting an utterly new flavor (that of durian fruit), and of receiving a sixth sense, unrelated to our first five, via a microchip implanted in the brain. She describes these experiences as "epistemically transformative," because one cannot have any idea what it will be like to have them until one does have them. In addition to new sense experiences, she classifies as "fundamentally inaccessible" the experiences of people who live a very different life from one's own: the white American businessman cannot understand what it is like to be a slave in the American South or an impoverished woman living in Ethiopia, nor can they understand what it is like to be him.

Though she acknowledges that one can "imagine something in the place of this experience you haven't had....this act of imagining isn't enough to let you *really* know what it is like to be an octopus, or to be a slave, or to be blind. You need to have the experience itself to know what it is really like." Paul thinks that just as we cannot know in advance what it will be like to have, e.g. a sixth sense, we cannot know in advance what it will be like to be a mother or vampire. Indeed, we have even more ignorance in the latter cases because they transform not only our sense-experience but also our value-experience. She cites feeling one's unborn baby kick, or feeling oneself move at super speed, as sense-experiences proper to motherhood or vampiricity, respectively. But of course the more dramatic changes are on the valuational front: becoming a mother means coming to care about one's child; becoming a vampire means, among other things, becoming indifferent to human suffering.

Paul laments the fact that a wall of ignorance deprives such agents of the knowledge that they most desperately need: "So, in many ways, large and small, as we live our lives, we find ourselves confronted with a brute fact about how little we can know about our futures, just when it is most important to us that we do know." (p.4) Presumably, the guiding example of a choice to become a vampire is supposed to emphasize the alienness of the new experience into which she takes the subject of transformative choices to leap.

One might object that Paul has overstated her case: surely inexperienced agents can know

something of the relevant experience in advance. We can see this by considering that one inexperienced agent might know more than another. If you've never seen any vampire movies, I know (at least a little) more about what it is like to be a vampire than you do. Someone who has had (any) new taste experiences has *some* conception of what it will be like to taste durian fruit—for instance, that the experience will have the character of a taste experience and not a hearing or seeing experience. Living with a blindfold for an extended period is not the same as being blind, but someone preparing to work with blind people might reasonably hope that it will give her an inkling of what it is like. Someone who has had a mother knows something of motherhood, and someone who has spent substantial amounts of time taking care of children might know a bit more. If you're a historian who has spent her life reading first and secondhand accounts of slavery (or Ethiopian villagers), you probably have a better conception of what that is like than I do²⁸.

In response, Paul might acknowledge that some of the inexperienced know *something* of what it will be like, but they do not know enough to warrant taking what they know into account. Her thought is that what we have is *too far* from firsthand experience to figure rationally in one's transformative choice. What are the standards for assessing whether someone's conception of what it will be like is close enough to what it will actually be like for her to be warranted in using that conception to guide her to that goal? Paul's approach to this question remains within decision-theoretic confines: she imagines that the way in which such a thought would have to figure is by offering you a basis for assessing the value of, e.g., being a mother. She presupposes that the rational transition to becoming a mother involves a comparison between (the experiential values of) motherhood and non-motherhood.

I agree with Paul insofar as her aim is limited to that of cautioning us against trying to integrate any conception of what it will be like to be a mother into a cost-benefit picture of the kind that suits medium-sized decisions. If I'm deciding whether or not to buy a new car, I might make a list of the advantages of having that car, as compared with the advantages of keeping my old car. It is possible for me to compare what it will be like for me to have a new car to what it will be like for me to have some extra cash. Whereas someone who is, e.g., indifferent to parenthood in the manner of Ullmann-Margalit's Old Person cannot 'weigh' her value-perspective against that of a person for whom family life is his *raison d'être*.

I disagree with Paul insofar as she concludes that thinking about what it will be like to be a mother is the mark of an irrational transition to motherhood. She suggests that instead of weighing motherhood against non-motherhood, I weigh my desire for stasis against my desire for change. I have explained why being leap-prone (pre-leap!) cannot carry enough rational weight to justify a substantial leap. (In short: because it would justify leaping anywhere.) I

²⁸ Paul takes her bearings from Frank Jackson's Mary, who does not come any closer to learning what red looks like, in her black and white room, by acquiring the totality of scientific knowledge about color and vision. But this does not show that there is no way to approximate the experience of seeing red: for instance, Mary knows more about what red looks like than blind Mary, who cannot so much as see the difference between black and white. Or suppose Mary leaves her room into a colorful room with no red in it. When she enters that room, she learns what it is like to see a color other than black and white. That puts her *closer* to the experience of seeing red than she was earlier. Indeed, with some creativity one can articulate this process of 'coming closer' into many small steps. Consider, for instance, these two steps between black and white Mary, on the one hand, and red-seeing Mary, on the other: (1) a Mary who has seen only cool colors (blue, green, etc.); (2) a Mary who has seen warm colors (orange, pink) but not red.

now propose that we give up on the idea of weighing or calculating or deliberating or comparing our way to motherhood.

I think that the rational agent both should and does think about what it will be like for her to be a mother, to live in France, to love music. In fact, aspiration is nothing other than the process of gradually working oneself all the way into that thought, to the point where what one thinks it is like is what it is in fact like. Without disagreeing with Paul that there is *some* sense in which what she calls transformative experiences are “fundamentally inaccessible” to those who do not have them, I am nonetheless going to set out to describe a path of rational access. We can think and act and feel and, more generally, *work* our way into motherhood. We can rationally bring ourselves to see things differently.

How is it possible to rationally proceed *into* a new point of view? First, we should note that it isn't always possible. We do not all have the opportunity to have children (or get married, become Israeli pioneers, etc.). But suppose someone *does* have the opportunity—let us simply assume the relevant external conditions are met. How does the change happen? On Paul's picture, thinking like a mother is something that happens to me as a result of an experience I decide to undergo. She models the mother case on the vampire case. When I become a vampire, I first decide to be bitten, and then the bite transforms me. After I am bitten, I will find myself with typically vampiric thoughts, habits and values: a love of spiders, excellent fashion sense and cold indifference toward humans. Likewise, in the mother case, I decide to have unprotected sex, and (we suppose) that results in pregnancy. But this is where the differences start. Being pregnant is quite different from being a mother. Some pregnant women have no intention of becoming mothers, and even those who do might struggle to see themselves that way. Seeing a positive pregnancy test can certainly be an emotional and shocking experience, but it does not magically endow one with the values, habits and feelings of parenthood²⁹. Those who struggle over whether to become parents—which is to say, anyone faced with what Paul calls a ‘transformative choice’—view parenthood not primarily as a biological category, but as an ethical one. This is why an adoptive mother or father is a mother or father in the unqualified sense. A parent is someone who raises a child, where childraising is an ethical process that includes nourishment, protection, education and love.

Unlike becoming a vampire, becoming a mother is not (just) something that happens to you—it is not even something you decide to have happen to you. It is something that you do. And it is not something you can do in a moment. We should not confuse what Ullmann-Margalit calls the “irreversibility” of some decisions—such as the decision to have unprotected sex, or leave one's homeland for Israel—with the idea that the transformation begins and ends at the moment at which one makes that decision. As Ullmann-Margalit acknowledges, there is one sense in which no decision is reversible. When one steps on to the boat or has unprotected sex, one takes a single step down a long path. The decision might be called ‘irreversible’ in that making it means that *not* taking further steps down that aspirational path will be very costly. I will discuss the implications of such costliness more below, but for now I want to point out that it does not follow, from the fact that avoiding becoming X-ish will be very costly to me, that I am already X-ish. Committing oneself to doing a kind of work is not the same thing as having done

²⁹ The same goes for any of the stages on the way to parenthood: hearing the heartbeat, feeling the baby kick, holding the newborn, giving him his first bath, etc.

it.

The decision to have unprotected sex is neither the end nor the beginning of the process of becoming a mother. A woman is likely to have started giving thought to the question of motherhood long before it presented itself as a real opportunity. Her thinking cannot end with that positive pregnancy test either, since she will have to use it to guide herself through the many forms of motherhood. Being a mother to an unborn fetus is a very different matter from being a mother to a teenager. We spend a long time becoming mothers.

Can a long transformative process of this sort be rational? Consider Ullmann-Margalit's distinction between opting and drifting. She observes that sometimes a person's character and values change ("drift") in incremental steps, so that one can see only in retrospect the magnitude of the change that those steps have added up to. She cites cases of people who slip, through a series of small and innocuous-seeming decisions, into criminal activity or marital infidelity. Ullmann-Margalit observes that we are apt to deceive ourselves into drifting through an opting situation by hiding from ourselves the magnitude of the change we are undergoing:

It is possible that from an outside-spectator's point of view the real nature of the actor's decisions is clear. It is possible for a person to proceed as a drifter while an informed spectator would judge that the person's situation is one of opting. When this happens, I think that we can view the actor as engaged in self-deception. The actor may be ignoring aspects of his or her decision situation, which reveal it for what it is: a first commitment leading down a core-transforming, irreversible road. (p.170)

Ullmann-Margalit's discussion of drifting suggests a suspicion of long transformative processes. Perhaps she would be inclined to view the claim that a long process is essential to someone's transformation into wife, mother or doctor, as evidence that the person is self-deceptively drifting as a way of avoiding opting. She cites sociological research that points to such a conclusion: "Important life decisions are sometimes incremental in nature, the end product of a series of small decisions that progressively commit the person to one particular course of action. A stepwise increase in commitment can end up locking the person into a career or marriage without his ever having made a definite decision about it."³⁰ The fact that Ullmann-Margalit's final discussion of "the place where opting and picking meet" takes an explicitly existentialist turn moves me to hear a Sartrean subtext in her discussion of drifting: (most?) human beings cannot handle the magnitude of the decisions that sometimes face them; they drift instead of opting, having abdicated their freedom by deceiving themselves into thinking they have no choice. These claims are at most implied by Ullmann-Margalit's suggestive final comments³¹, but I wanted to make them explicit in order to address them here.

³⁰ Janis and Mann, cited in Ullmann-Margalit 2003 p.170.

³¹ In support of ascribing these implication to Ullmann-Margalit, I observe that her other paper on this topic, written together with Sidney Morgenbesser 30 years earlier, ends with similar approving reference to existentialism as "a philosophical literature...suggesting that at the very deepest level of selection, involving the ultimate and most significant alternatives confronted by man, there can only be picking, there being no possibility of a reasoned choice." That paper ends by gesturing at the position she was to articulate later: "given our beliefs and utilities, we pick or we choose as the case may be but as to our utilities or values themselves, to the extent that they can be thought to be selected at all, they can only be picked." (p.783)

But it is not true that the time spent becoming a mother, or a wife, or a doctor is (necessarily) time spent drifting. Becoming something little by little is not only something that happens to us, it is (at the risk of repeating myself) also, sometimes, something we do. Recall Paul's separation between the event of deciding to become a vampire, and the event of becoming one. On Paul's conception of transformative choice, agency ends just before one becomes a vampire. I am trying to articulate the form of agency that crosses this boundary: the agency of becoming. Ullmann-Margalit might ask why, if you know you want to be a mother (or vampire, or music-lover), you don't go all the way at once. The answer is that you don't know, not fully, since you're still learning. Becoming a mother is hard, and it takes work, and work takes time. We may only be able to do a little of it at once. Changing oneself is, as they say, a process.

Ullmann-Margalit and Paul may object that I have changed the topic on them. They wanted to know *how we rationally decide* to become mothers or pioneers. I have responded by gesturing at the existence of a rational, temporally extended process by which someone becomes more and more of a mother or a pioneer. Their arguments analyze a moment of decision in which someone, as Ullmann-Margalit would say, "opts" to board the boat to Israel or to throw away her birth-control pills. My claim is that those moments are part of a longer process; and it is the process that is rational. I find their focus on climactic, life defining choice distorting. For instance, Ullmann-Margalit asks us to "think of a young talented person who faces a choice between a career as a concert pianist and as a nuclear physicist." (p.160) But there is reason to doubt that anyone has ever faced this choice. One cannot simply "choose" to become a nuclear physicist or a pianist, regardless of one's innate talent. These careers require singleminded dedication; they are a product of work, rather than choice. Nonetheless, Paul and Ullmann-Margalit might insist, we do sometimes need to decide whether or not to board the boat; whether or not to trash the pills; etc. How do we make this decision rationally? That was their question, and I haven't told them how to answer it.

Ullmann-Margalit notes that we if we look at how people actually make these decisions, we will be disappointed:

"one would expect the opters to take extra time and care in amassing relevant information as their evidence base, to exercise extra caution in assessing the alternatives open to them....In short, one would expect an act of opting to be an exemplary candidate for the ideal rational-choice explanations just delineated....There is some evidence that the attitude of people toward their big decisions is quite the opposite of the one that we might expect. That is to say, evidence³² seems to suggest that people are in fact more casual and cavalier in the way they handle their big decisions than in the way they handle their ordinary decisions." (p.165)

Ullmann-Margalit's guiding question is the normative question of how someone might handle a big decision rationally, and so she quickly sets aside her own surprise at the social-psychological evidence for the conclusion that we do not, as a matter of fact, do so rationally. But I think this answer to the descriptive question should give us pause. Whenever we find ourselves concluding that some fundamental human activity is characteristically irrational or arational, we need to consider whether it is possible that we haven't quite grasped the phenomenon. In this case, I think the mistake is one of overly narrow focus. If you look at *the*

³² Ullmann-Margalit supports this claim with a series of references to social-psychology literature I will not reproduce here (see her footnotes 7 and 8).

moment in which someone decides on college A instead of college B, or decides to get married or to buy a house, you've ignored the context necessary for getting the relevant practical rationality into view.

Going to college is indeed a momentous event, one the high school student has been looking forward to for years; and when she gets to college she will continue to struggle to understand what it means to be a college student. Choosing college A instead of college B may well be a blip in this process. Even though she will have different experiences and become a very different person at the two different places, she is unlikely to be able to anticipate those differences in advance. From the point of view of her aspiration to learn and grow in college, it is quite likely that the colleges she is in a position to decide between present a virtually identical face. And thus if she does not trouble herself much over the choice, this is not because she adopts a "casual and cavalier" attitude to her own education. She is likely to care deeply about the person she will become, what she will learn, how she will grow. But this caring is not expressed in the moment of deciding between the two colleges.

Deciding between colleges may well be picking, but this does not show that "at the very deepest level of selection, involving the ultimate and most significant alternatives confronted by man, there can only be picking." (Ullmann-Margalit and Morgenbesser p.783) For the deepest level is in fact one level up: going to college at all, rather than going to *this* college. It may, perhaps generally, be said that 'small choices' for which picking is appropriate are cases in which one has adopted too fine-grained a description of the decision scenario. The shopper decides to buy cereal instead of oatmeal³³, he does not decide to buy *this box* of cereal instead of that one. He simply *does* buy this box instead of that one, without making a decision on the matter. If picking is not deciding, then decision theorists do not fail us by not offering a procedure for picking rationally. I propose that we can defend decision theory in a very similar way with reference to what Ullmann-Margalit calls large decisions: they are not decisions. The rationality of transformative experience is not visible to one who looks at only a moment in the process. This is true even for certain 'climactic' moments, discussed in more detail below, such as the moment one opts to board a boat or throw away the birth control pills or accept college A over college B. If you want to see the rationality of transformative experiences, you will have to look at them in a different way, as part of a longer story. But that entails that the rationality of transformative experience is not a matter of choice. The decision theorist needn't offer us any account of how we rationally become mothers or pioneers, since that topic lies outside her purview. The person whose job it is to describe that rational process is the theorist of aspiration.

Recall Old Person, who is (at t1) committed to the exciting single life, but who ends up (at t3) as a boring parent, New Person. Suppose you are watching a movie about Old Person which slowly presents highlights of his transformation, via experiences with his friends' children, private moments of reflection, developments in his relationship with his partner or his own parents, etc., to the moment (t2) where he says "let's go for it,"; the movie continues to present his slow and sometimes painful growth into his role as a parent until the point where he says (t3) he is happier than he has ever been. That is likely to be a bad movie, because the change

³³ Or to go shopping instead of watching TV. We will have to look to the shopper's attitude to pick out the right level of description. (see fn. 19)

described is *so* intelligible as to be clichéd. We would not find ourselves forced to classify the decision at t2 among acts of the will that are “nihilist, absurd, or leaps (of faith).” (p.172) It is not like Meursault’s or Raskolnikov’s decision to become wanted criminals by killing some random person. Ullmann-Margalit considers it difficult to judge the rationality of his transition because we cannot choose between adopting old person’s t1 perspective on the value of parenting, and his t3 perspective on the value of parenting. But this leaves out the fact that at some point after t1 but before t2, Old Person himself starts to try to adopt the t3 perspective. By t2, he is trying to see the world, more and more, from a point of view that appreciates the life of a parent. The changes in value that he undergoes are responses to his attempt to value differently. His transition from Old person to New person will look absurd only so long as we leave out the fact that during it, Old Person is trying to become New Person.

VI. Deciding to Become Someone New?

Paul and Ullmann-Margalit ask after the agent’s reasons for becoming a mother, wife, Zionist, college student etc. I have been arguing that they have not found these reasons because they have looked for them in the wrong place, expecting these reasons to surface at some moment of decision rather than in the years-long process of learning, growth and value-change. The reasons of the aspirant are the reasons she responds to throughout the temporally extended process of becoming a mother. I will describe these reasons in the next chapter. Before doing so, however, I want to devote some attention to the climactic decisions that Paul and Ullmann-Margalit have taken to be so important for understanding the rationality of transformative experience.

It is natural to say that at the moment at which an agent is poised to board the boat for Israel, throw away her birth control pills, or submit her college application, she is “deciding whether or not to become” a Zionist, mother or college student. This locution is itself harmless, so long as we use it in a way that is compatible with not taking the decision as the rational basis of her transformation. In order to see what function it does have, let me first underscore a point made above: the decision marks neither the beginning nor the end of aspiration.

The people I described wouldn’t be hesitating or wavering over their decision if they had not *already* begun the project of becoming the people in question. They are trying to decide whether or not to *continue* on an aspirational path they have already traveled some ways down. They hesitate before a particularly costly such step. We can see the fact that they have already come a certain distance in the fact that the prospect of exiting is experienced as (at least somewhat) painful: someone who decides not to continue will experience that decision as involving some kind of loss, and this means she must already have some grip on the value³⁴. Thus it is not the case that climactic decisions represent the agent’s first acquaintance with some value. In this sense, they are not “deciding to become” the people in question.

Nor will the decision in question complete their transformation. Getting on the boat, throwing the way the pills, or sending in the application is but a step on the path to becoming someone. These steps are distinctive because the stakes are high. We will have sacrificed many things for

³⁴ Paul takes this experience of loss to be illusory, on the grounds that the agent cannot, at this point, have any sense of what she is missing. I argue to the contrary in the conclusion of this book [PAGE REF].

what may turn out to be nothing, or worse than nothing. With some aspirational projects, our failure to succeed at them will simply mean that we turn our lives in a different direction; by contrast, in these high-stakes cases, the penalty for failure may be misery or even death. But this is just to say that the choice itself does not ensure success: for precisely what one fears is that one will *not* see one's way to finding meaning in motherhood, Zionism, or college education. Notice, also, that a decision to turn away from motherhood, Israel or college does not extinguish the aspirational impulse in that direction. We cannot decide not to aspire any more than we can decide to aspire; and this is because we cannot change our own evaluative condition at will. The agent who steps away from the boat, continues taking the pills, or doesn't submit the application will live, at least for some time, with the value residue of what she has lost. In chapter III I discuss such agents, who fall among the class of people I describe as having a fractured evaluative perspective.

These climactic decisions showcase a distinctive kind of practical problem that aspirants find themselves in: they sometimes have to make a decision as to the value of continuing on their aspirational path, while recognizing they will only really know that value if they do continue. When we have the power, we try to shield our children or friends from being 'forced to decide.' We do not want college students to be weighing the value of a college education against the hardship of the loans they have to take in order to finance it; we feel for someone who hesitates over motherhood because it entails a sacrifice in career or personal fulfillment; we empathize with a person who has been offered a promising job abroad when she has just begun a romantic relationship. We are moved by the plight of the Zionist pioneer who must weight her duties to her family against the new life she feels she could have in Israel. The protective impulse we feel when we consider these stories is not only a desire for an ideal world in which people can have everything they want. It is, I believe, an impulse to shield someone from having to assess a value she is only beginning to get to know. The problem is akin to that of giving someone a math test in her first week of class. These choices force someone, who is still reasoning toward desire, to reason from desire.

These decisions are difficult, but not as difficult as Paul and Ullmann-Margalit take them to be. For the agent who is reasoning toward desire or value is, gradually, acquiring the desire or value. And she can therefore reason from whatever grip on that desire or value she has already attained. It is true that, aware that her grip is incomplete, she will not be confident comparing the value against *major* costs that its pursuit will incur. She sees she would be in a better position to make those comparisons later on, when she is more acquainted with the value. But, as luck would have it, she must make them now. This bit of bad luck is not distinctive of aspiration—we often have to make big decisions without having all the information we would like, and, specifically, we often have to give something up without knowing just how great a value we are thereby losing. And so aspirants, like many other agents, will simply have to ask for advice, take risks, and hope for the best. Paul and Ullmann-Margalit are wrong to assume that such agents have *no* basis for making such choices; but more importantly, they go wrong in taking their bearings on the aspirant's understanding of what she is doing overall from these moments of high-stakes choice. The fact that aspirants tend to face such forced choices reveals less about the rationality of the aspirational process overall than it does about the unfortunateness (and in some cases injustice) of certain social arrangements.

I have said that the climactic decisions described by Paul and Ullmann-Margalit presuppose an agent who has already begun aspiring. One might wonder when or how such agents got started. Before we ever have to make a choice such as whether to get married, go to college, or have children, we undergo extensive education on the value of these activities at the hands of our family, friends, teachers and acquaintances. As I will argue in chapter 6, aspirational agency is distinctively dependent on such environmental support. One important element of such dependence is dependence on others for orchestrating original contact with values that will eventually become objects of aspirational pursuit. Someone who has never been exposed to forms of caregiving that would make it look appealing is unlikely to have even a slight inclination toward motherhood. Our initial contact with the value is a germ whose development into full-blown appreciation is the self-creation process I describe in chapter 5. But that is not to say that the agent creates the germ. I will explain, there, why we should not think her agential control over the later stages is dependent on her having much, or any agential control over the early ones. I mention these upcoming arguments only as a way of gesturing at an answer to a question that may arise about the origins of aspiration: we do not exercise agency over that event. Aspiration begins before we are in a position to exercise agential control over our relation to the value: it gets started, but we don't start it. This is not only because we are too young to do so, but also because, for reasons well brought out by Paul, we could not have any reason to bring ourselves into contact with something for which we have *no* preference, appreciation, value, etc. Someone could behave in a way that resulted in such contact, but she would not be acting in order to come into contact with it.

* Chapter 2: Proleptic Reasons

The teacher of a music appreciation class is frustrated with those students who are taking her class, as she puts it, “for the wrong reasons.” In her view, the class offers students access to the intrinsic value of music. Students who are taking it for “the right reason” will be taking it for this reason. But only those who already appreciate music appreciate musical appreciation. Or, at any rate, only they appreciate it correctly, for the reason for which (she believes) one should appreciate it, namely, intrinsic musical value. The problem is that if the intrinsic value of music is a reason you respond to, you don’t need to take her class. You already appreciate music.

She wants students in the class who care about music. But she’s supposed to be teaching them to care about music. Is she being unreasonable? The problem does not go away once we admit of degrees or kinds of caring—it does not help to characterize her job as that of getting people who care a little (or who care in this way) to care more (or in that way). So long as someone enters the class satisfied with his level or type of music appreciation, whatever that may be, the teacher will impugn his motives, whatever they may be. The teacher is looking for students who want to care about music more than, or in a different way than, they currently do. But, again, she doesn’t want them to want this for some extra-musical reason. So it seems that what she wants is for them to respond to musical value exactly to the extent that they’re not yet able to.

This is a paradoxical way of stating an ordinary demand for the kind of reason that is my topic. It is possible to have an inkling of a value that you do not fully grasp, to feel the defect in your valuation, and to work toward improvement. The reason for doing that work is provided by the value in question, but the defect in your grasp of that value also shapes the character of the activity it motivates. For consider what kind of thinking motivates a good student to force herself to listen to a symphony when she feels herself dozing off: she reminds herself that her grade, and the teacher’s opinion of her, depends on the essay she will write about this piece; or she promises herself a chocolate treat when she reaches the end; or she’s in a glass-walled listening room of the library, conscious of other students’ eyes on her; or perhaps she conjures up a romanticized image of her future, musical self, such as that of entering the warm light of a concert hall on a snowy evening. Someone who already valued music wouldn’t need to motivate herself in any of these ways. She wouldn’t have to try so hard.

The paradox arises from a dilemma concerning two kinds of reasons a potential student of such a class could have for taking it. There is, first, the intra-musical reason, the having of which seems to mark the fact that the class has come to a successful *close*. There is, second, any extra-musical reason, the recourse to which seems to condemn someone to subordinating the value of music to what the teacher would call “an ulterior motive.” In the first case, the reason is not the reason of a *student*, in the second case, it’s the reason of (what the teacher would call) a *bad* student. I will argue that this dilemma is specious, because there is an agent—the good student—who manages to combine extra- and intra-musical reasoning. Like the music-lover she will become, she is genuinely oriented toward the intrinsic value of music. For instance, if offered some way of attaining good grades, chocolate treats, etc. without coming to appreciate music, she would reject it. And yet grades and chocolates are integral to the rational explanation of her action of listening to music: she would be asleep without them. ‘Bad’

reasons are how she moves herself forwards, all the while seeing them *as* bad, which is to say, as placeholders for the ‘real’ reason.

One characteristic of someone motivated by these complex reasons, by contrast with the simpler reasons of the bad student, on the one hand, and the established music lover, on the other hand, is some form of embarrassment or dissatisfaction with oneself. She is pained to admit, to herself or others, that she can only ‘get herself’ to listen to music through those various stratagems. She sees her own motivational condition as in some way imperfectly responsive to the reasons that are out there. Nonetheless, her self-acknowledged rational imperfection does not amount to akrasia, wrongdoing, error or, more generally, any form of irrationality. Something can be imperfect in virtue of being undeveloped or immature, as distinct from wrong or bad or erroneous. (There is something wrong with a lion that cannot run fast, but there is nothing wrong with a baby lion that cannot run fast.) When the good student of music actively tries to listen, she exhibits not irrationality but a distinctive form of rationality.

Her rationality is not, however, of the familiar, clear-eyed kind. Anscombe’s Intention placed the ability to answer the “why?” question at the heart of philosophical discussions of agency. The agent who can give an account of what is to be gotten out of what she is doing grasps the value of what she will (if successful) achieve through her action. Her answer to the “why?” question might not satisfy every interlocutor, but it is at least satisfying to the agent herself: she takes herself to know why she is doing whatever it is she takes herself to be doing. Of course, not every agent will be able to satisfy herself in this way: some agents are not paying attention to what they are doing, or are being impulsive,³⁵ or experience a moment of forgetfulness, or have simply failed to think things through sufficiently. In some of these cases, the agent’s behavior is arational, since her ignorance is profound enough to disqualify her from acting intentionally; in other cases, her action is intentional, but irrational. The good student of music likewise fails to be able to articulate, to her own satisfaction, what she expects to get out of her music class. In her case, however, this marks neither the absence of intentionality nor the absence of rationality.

If an agent finds her own answer to the “why?” question satisfying, she must ascribe to herself a certain knowledge of value. Such an agent takes herself to know both that some form of value is on offer, and that it is one she herself does or will enjoy, appreciate, or find meaningful. And such a person is often correct—agents often do have such knowledge. How did they acquire it? Since knowledge of value is itself valuable, it stands to reason that one way we acquire such knowledge is the way we acquire many other valuable things: by acting in order to bring about that acquisition. The problem is that unless one is equipped with an ulterior motive, the value of knowledge of some value is not a different value from that value itself³⁶. Therefore, those seeking to acquire the knowledge cannot take themselves to know why they are doing so. And yet—I will argue—it is a fact of life that people act not only from, but also, at other times, for the sake of acquiring, knowledge of value.

³⁵ Some, but not all, impulsive agents will take themselves to fail in respect to the “why” question. “Just because I feel like it” might strike one agent as a perfectly good answer, and another as no answer at all.

³⁶ As in Frankfurt’s (1971) example of the doctor who treats drug addicts: he wants to understand the appeal of drug addiction without actually wanting to become addicted.

If those actions are to be rational, then rationality cannot require accurate foreknowledge of the good your rational action will bring you. Thus I will defend the view that you can act rationally even if your antecedent conception of the good for the sake of which you act is not quite on target—and you know that. In these cases, you do not demand that the end result of your agency match a preconceived schema, for you hope, eventually, to get more out of what you are doing than you can yet conceive of. I'll call this kind of rationality, "proleptic." The word 'proleptic' refers, usually in a grammatical context, to something taken in advance of its rightful place. I appropriate it for moral psychology on the model of Margaret Little's phrase 'proleptic engagement,' (p.342) by which she refers to an interaction with a child in which we treat her as though she were the adult we want her to become³⁷. Proleptic reasons are provisional in a way that reflects the provisionality of the agent's own knowledge and development: her inchoate, anticipatory and indirect grasp of some good she is trying to know better. Proleptic reasons allow you to be rational even when you know that your reasons aren't exactly the right ones.

A reason for action is a consideration in favor of acting in some way; if the agent in fact acts on the reason, she will be able to offer that reason as an explanation of why she so acted³⁸. Sometimes we do something for more than one reason: I went to the store in order to get milk and for the exercise. Proleptic reasons are double in a more fundamental way. The good music appreciation student is listening to the symphony assigned for her class because music is intrinsically valuable, and because she wants a good grade. If she merely cited the first as her reason, she would be pretending to a greater love of music than she currently has; if she merely cited the second, she would be incorrectly assimilating herself to the bad student. But her motivational condition is also not one in which she has merely added the first reason to the second, because that situation would describe a music lover who is (strangely) taking a music appreciation class. The fact that music is intrinsically valuable and the fact that she wants a good grade somehow combine into *one* reason that motivates her to listen. The reason on which she acts has two faces: a proximate face that reflects the kinds of things that appeal to the person she is now, and a distal one that reflects the character and motivation of the person she is trying to be. Her reason is double because she herself is in transition.

I will show, by generalizing the paradox described above, that it is not only the rarified context of music education that calls for a proleptic analysis. I argue that we must acknowledge the reality of proleptic reasons, else we be forced to classify as irrational a large swath of human agency—agency that is purposive, self-conscious, intelligent, truth-sensitive, and constitutes a kind of building block of or prelude to everything else that we do. I end with a discussion of the currently dominant moral psychological thesis that what practical reasons we have depends on what desires we have. I consider a few variants of such "internalism," as it is called, and argue that none of them can, as they stand, make room for the existence of proleptic reasons.

³⁷ Likewise Bernard Williams (1995) speaks of a 'proleptic mechanism' by which he takes at least some instances of blame to function. Williams asserts that a blamer's pronouncement that the blamee "ought to have ϕ -ed" can serve not as a description of the blamee's current set of reasons, but rather as a way of both anticipating and bringing about the future state of affairs in which the blamee will be in a position to be motivated by the reasons now being ascribed to him.

³⁸ The reason in question is at least a motivating reason, and it may also be a normative one. For discussion of that distinction, see below, fn. 45.

I. Large Scale Transformative Pursuits

In the previous chapter, we discussed momentous life-changes under the heading of ‘big choices’ or ‘transformative experiences.’ I ended by arguing that what Paul and Ullmann-Margalit characterize as decisions or experiences are better understood as temporally extended, agentially directed processes. Let us, therefore, adopt the phrase ‘large-scale transformative pursuits’ to describe such significant life changes as: attending college, moving to a foreign country, adopting a child, becoming a painter or a philosopher or a police officer, achieving distinction in athletics or chess or music, becoming a sports fan, an opera lover or a gourmet, befriending or marrying or mentoring someone, etc. The features uniting this class of pursuits³⁹ are that they change what one cares about, and that they change it in some substantial way. They typically require years of sustained effort, both in the form of preparation and in the form of the work attending the completed state. They are both transformative and large in scale.

We have labeled the class of pursuits that involve a small-scale change in what one cares about “self-cultivation.” (See ch. 1, III) If I inculcate in myself preferences whose value can be fully cashed out in terms of my current preferences, it is easy to rationalize my action. When a pursuit is large in scale without being transformative, I will describe the agent as ambitious. (See below, Chapter 6 part III.) Wanting to cure cancer, make a million dollars, or win the Nobel Prize can, if understood in a sufficiently narrow way, count as large-scale nontransformative pursuits. The key feature of such cases that allows one to classify them as ambitious is the presumption of value-stasis: one needn’t oneself undergo a value change in order to succeed at the project.

Transformative ends are recognized as such not only by those who have succeeded in attaining them, but also by those who are on their way: one can see in advance that one cannot see in advance all of what is good about parenthood, or friendship, or scuba-diving, or immigrating to another country. Transformative pursuits aim at values, the appreciation of which is connected to the performance of the activity (or involvement in the relationship) in question. Indeed, this is because the pursuits themselves form a kind of value-education, gradually changing the agent into the kind of person who can appreciate the value of the activity or relationship or state of affairs that constitutes the end of the pursuit. In the course of becoming a teacher or a friend or a reader of ancient Greek, one learns to appreciate the values that are distinctive of teaching or friendship or reading ancient Greek.

But one does not fully appreciate them until one is at, or close to, the end of the process of transformation. For it is the end-state (teaching, parenting, translating) that offers up the actual engagement with the value on which any full appreciation of it must be conditioned. The joys of teaching are best known to teachers. Everyone goes to college “to become educated,” but until I am educated I do not really know what an education is, or why it is important. I may say I am studying chemistry in order to understand the ‘structure of matter,’ but only a chemist understands what it means for matter to have structure (or, indeed, what matter really is). For the rest of us, that phrase is likely to be backed by little more than an image of a tinker-toy “structure” to which a mental label such as ‘molecule’ is affixed.

³⁹ Though see fn. 19 for a caveat: this list represents what are, by and large, big changes in anyone’s life. I do not deny that there could be people for whom these changes barely register.

The problem posed by large-scale transformative pursuits is this: they require us to act on reasons that reflect a grasp of the value we are working so hard and so long to come into contact with, but we can know that value only once we have come into contact with it. And yet the cost of granting that such ends are pursued for no reason, or bad reasons, would be to restrict the scope of practical rationality very greatly. For most, if not all, of the experiences, forms of knowledge, ethical and intellectual traits, activities, achievements, and relationships that we value are such that the pursuit of them is both large in scale and transformative. It is true that even if we were forced to characterize the choices by which we move ourselves toward all of those ends as irrational, we could still rationalize engagement with the ends once achieved. But if this is all there is to practical rationality, we should be disappointed. For every rational choice to continue in some pursuit will be adventitiously predicated on a series of irrational choices to begin that pursuit. We should expect more from our reasons than maintenance of a mysteriously attained status-quo. I propose, therefore, to introduce a species of reasons to meet this expectation.

My music-appreciation example built in a demand, on the part of the teacher, that we not separate the rationalization of the pursuit from that of the end⁴⁰. This kind of demand is generally appropriate for large scale transformative pursuits. We do not want to understand them along the lines of someone who walks to the park for the exercise, but stays when she sees they're showing an outdoor movie. For in that case the agent was not, when walking, pursuing the end of seeing a movie. It is possible to rationalize both the walk and the movie-watching without rationalizing anything we could call the *pursuit* of the movie. By contrast, large scale transformational pursuits are characteristically aspirational: when the agent gets where she's going, she sees that she has what she was after all along.

II. Alternatives to Proleptic Reasons?

We ought to demand a rational account of how someone can work her way to the valuation characteristic of the various end-states to which she aspires. Satisfying this demand, I claim, means postulating a set of reasons—I've called them "proleptic reasons"—tailor-made to rationalize exactly these sorts of pursuits. By way of argument for this claim, let us survey alternate contenders, reviewing the kinds of factors we typically cite in explaining such behavior: a vague grasp of the value in question; a precise grasp of a value in close proximity to the value in question; reliance on the ethical testimony of a mentor or advisor figure; imaginative engagement in a pretense of being as one aspires to be; competitiveness; recourse to self-management techniques of (dis)incentivization. I'll argue, case by case, that vague reasons, approximating reasons, testimonial reasons, reasons of pretense, competitive reasons, and reasons of self-management rationalize in the right way only insofar as we help ourselves to a dedicated subset of each genus of reasons. It turns out that in order to rationalize aspirational agency, we must invoke not vague reasons but *proleptically vague* reasons, not testimonial reasons but *proleptically testimonial* reasons, etc. In the attempt to avoid proleptic rationality, we find ourselves ushering it in piecemeal, through the backdoor.

⁴⁰ David Schmidtz' (1994) maieutic reasons are intended to do the work I assign to proleptic reasons, but they fail this criterion: Schmidtz separates the rationalization of the end from the process by which we arrive at it.

1. Vague Reasons

Someone who has a ‘vague reason’ for ϕ -ing ϕ -es with only a vague idea of the value of ϕ -ing. It is certainly true that I have a vague idea of the value of all sorts of pursuits in which I am not currently engaged. For instance, I think there are many valuable careers I did not choose, many valuable hobbies I don’t pursue, many valuable books I’m not reading. One problem with such ideas is that they are often not very motivating. I don’t plan to read most of those books. Consider a *bad* student of music appreciation, one intent on merely going through the motions necessary for fulfilling a distribution requirement. He might happily grant that music-appreciation is a “good and valuable end.” He has a vague idea that music-appreciation is good. But that’s not enough to get him to do the homework, show up to class on time, study for the exam etc. A vague idea does not entail willingness to put in effort. So let us suppose that the vague idea is not *so* vague—in fact, let us posit that it suffices for motivation. There are many non-aspirational situations in which I have only a vague idea of the value I am motivated to get. I buy tickets to an opera I know I love, not knowing exactly what I will love about this production. Such an activity is not aspirational, because I’m satisfied with my vague idea. I don’t now feel the need to work to make up the difference between the vague idea I have now and the sharp one I will have later; I don’t experience that difference as a defect in my current state. I need only wait for the world and my interests to line up in such a way as to make it possible for me to do the enjoying or appreciating that I’m already fully capable of.

The aspirant’s idea of the goodness of her end is characterized by a distinctive kind of vagueness, one she experiences as defective and in need of remedy. She is not satisfied with her own conception of the end, and does not feel that arriving at the correct conception is simply a matter of waiting. She understands her aspirational activity as work she is doing toward grasping this end. So, while vague conceptions of value do help explain how aspiration is possible, it is equally true that the phenomenon of aspiration helps us understand a distinctive form of vagueness—a kind of ever-sharpening vagueness. Large scale transformation pursuits are done for those vague reasons that are *proleptically* vague.

2. Self-Management Reasons

My music student plans to reward herself with chocolate for getting through the symphony. I might make plans with a buddy to go running in the morning, so that she can hold me to my plan. Reasons of self-management show up whenever I am trying to get myself to do something that I think I should do but may feel insufficiently motivated to do. Some forms of self-management can be very mild, such as simply resolving to (not) do something. In all these cases, I find some way to add motivational backing to a given course of action. Notice, however, that such self-manipulation comes in two forms.

Suppose that Sue worries that she’ll be tempted to buy expensive holiday presents⁴¹ for her friends, despite her lack of funds. So she adopts one or more of such self-managing tactics as: choosing a thrifty friend as a shopping partner, leaving her credit card at home, resolving not to enter a certain expensive store. In the case I’m imagining, Sue does not see her temperamental generosity as problematic. She doesn’t have a systemic problem, she just happens to be very

⁴¹ I thank Kate Manne for the example, and for helping me to see its importance.

short of funds at the moment. Reasons of self-management are, in this kind of case, directed only at *behavior* on a given (or even a series) of occasions.

A different kind of holiday shopper might, by contrast, be engaged in a long-term struggle to curb her chronic overspending by learning to think less commercially about how to make herself and those around her happy. In that kind of case, self-management is directed primarily at changing how the agent thinks, values, and feels. The music student described above would presumably see it as quite problematic if, years hence, she were *still* motivating herself to listen with chocolate. Or consider the case of moving to a new country. I may, at first, have to ‘force’ myself into social situations. My hope is to thereby come to inhabit the new culture, language etc. in such a way as to become disposed to eagerly engage in such socializing. I aspire to make this new place my home. This second kind of self-management often goes along with a characteristically aspirational form of *practice*. In some cases, doing something over and over again changes the way I do it. And so by doing it, I hope to change my attitude toward it. Sometimes I manage myself precisely with the aim of managing myself less and less. And that is just to say: reasons of self-management, too, come in a proleptic variety.

3. Testimonial Reasons

We often invoke testimony to explain how someone’s rationally held beliefs can outstrip the cognitive resources that can strictly be called his own. There is some controversy over whether such testimony is possible in a moral context⁴², but it certainly seems possible to heed the practical advice of your elders and betters—even against your own instincts and inclinations. It is also true that advisors or mentors often, even typically, figure in large scale transformative pursuits. But the mentor’s role in the life of the aspirant is not an unproblematic one. Unlike in other testimonial contexts, the aspirant’s goal is nothing other than coming to see the value for herself. The fact that your role-model knows so much more than you that you are inclined to defer to her advice means that contact with her is a constant reminder of what you don’t have. You don’t aspire to do what she does, you aspire to do what she does in just the way she does it—namely, independently.

What would the music-appreciation teacher think of a student who takes her class on the advice of his music-loving mentor? I think the teacher would be satisfied with this reason to the extent that she felt the student wasn’t. I’m happy to take someone else’s word about the truth of many of my historical or scientific beliefs. I’m not, similarly, *happy* with my reliance on my mentor. The species of testimonial reasons that figure in aspiration are special in just the way that the vagueness of an aspirant’s conception of her end is special. The testimonial element in aspiration is of a distinctively degenerative kind: the present legitimacy and authority of the mentor’s voice is conditioned on, indeed, anticipates, its gradual evanescence. And in characterizing this curious species of testimony we have, once again, helped ourselves to a dedicated, aspirational species of the genus in question.

⁴² See Wiland (2014) and McGrath (2011).

4. Reasons of Competition

Many large scale transformative pursuits are, at some point or other, fueled by a desire to position oneself at the top of some group of people engaged in a similar pursuit. Wanting to be better than others at something is a very powerful motive. The mathematician G.H. Hardy writes that he initially “thought of mathematics in terms of examinations and scholarships: I wanted to beat other boys, and this seemed to be the way in which I could do so most decisively.” (*A Mathematician’s Apology*, p.46) We frequently encounter such competitiveness in athletic, musical, intellectual and artistic pursuits. People even get competitive about their hobbies. But there are—again—two kinds of competitiveness.

In one kind of case, I compete in order to display my excellence or submit it for assessment. So: I would like my excellence to be praised, celebrated, renowned to others. Or I would like to know how good I am, perhaps to be reassured that I really am as good as people say I am. Competition can be a way of gauging one’s excellence, by measuring it against the excellence of others, or flaunting it, by demonstrating its superiority to the excellence of others. Such flaunting can itself spring from a variety of motives—for instance, I might want to flaunt my excellence as a physicist in order to inspire other young women to become physicists. Whatever the ultimate motive, competition of this kind is characterized by a desire to make known to others or to myself a virtue that I *already* have.

In another kind of case, the point of competition is to allow me to strive for excellence in an open-ended way. The thought of being better than the people around me is a powerful motivator for making something of myself when I don’t know exactly *what* it is I want to make of myself. Hardy recounts:

“I found at once, when I came to Cambridge, that a Fellowship implied ‘original work’, but it was a long time before I formed any definite idea of research. I had of course found at school, as every future mathematician does, that I could often do things much better than my teachers; and even at Cambridge, I found, though naturally much less frequently, that I could sometimes do things better than the College lecturers. But I was really quite ignorant, even when I took the Tripos, of the subjects on which I have spent the rest of my life; and I still thought of mathematics as essentially a ‘competitive’ subject.” (p.47)

If the motivations driving Hardy to become one of the twentieth century’s greatest mathematicians were competitive in nature, this competitiveness must have been of a singularly consuming kind. In this kind of case, competitiveness is a way of holding open a door for the person I’m trying to become. I’m competing *in order to become* excellent, rather than in order to show that I already am. When the prize arrives it turns out to be not what I really wanted; I am already preparing for the next competition. The value for the sake of which I compete is not one on which I have a good grip. I compete for the sake of a future or anticipated value that I, as of now, only incompletely understand. This form of competitiveness is *proleptic* competitiveness.

5. Reasons of Pretense

David Velleman has proposed that we emulate ideals by *pretending* to satisfy them. He offers as an example of pretense his own experiences of mock-aggression in his martial arts class. He then analyzes a case of quitting smoking as one in which the subject pretends to be a non-smoker and then gets “carried away” (p.100 *et passim*) with the pretense. Velleman

acknowledges that, on his conception of it, such behavior is somewhat irrational: “when a smoker draws on an ideal for motivation to quit, his behavior is in some respects irrational.” (p.101) He characterizes such agents as “hav[ing] reasons to make themselves temporarily irrational.” Velleman seems to think that the irrationality in question is only of a harmless, temporary kind. I find it to be neither harmless or temporary. The whole idea of such an account is to sever someone’s ‘outer’ reasons for adopting the pretense from the reasons as they appear to him once he’s inside it. Velleman’s thought is that the agent thereby makes a new set of reasons available to himself, which he can leverage into personal change. But once one adopts an account of this kind, one cannot rely on the rationality of the outer reasons to vouchsafe that of the inner ones. Consider that one can have all sorts of reasons for ‘pretending’ to be some way—someone can pay me money, I can do it on a lark, I can be an actor in a play. If I get “carried away” and fail to snap out of it, I seem to exhibit some kind of mental illness. I’ve become trapped inside my own game. Velleman offers no principled reason why we should not understand the smoker, and emulation in general, as (possibly⁴³) luckier victims of the same deep and permanent irrationality.

In aspirational cases, the failure to shed the pretense is salutary rather than pathological. But this is connected to the fact that it is not mere pretense. When I pretend or engage in make-believe, I close my eyes to the world around me, sometimes literally, the better to imagine a world that isn’t actually there. It is crucial to my willingness to engage in such activity that I see it as temporary. Large scale transformative projects—including that of quitting smoking—are not like this. If I aspire to become a nonsmoker, I am not pretending to already be one. Rather, I want to come to see the world in the way in which a nonsmoker does, because I think that is the right way to see things⁴⁴. I’m not closing my eyes, I’m fighting to open them and to *keep* them open. Velleman’s conception of aspiration corresponds to Iris Murdoch’s description of humanity in general: “man is the creature who makes pictures of himself and then comes to resemble the picture.” (p.252) I think the aspirant makes pictures of himself *in order to* resemble the picture.

Pretending is different from trying, but I don’t want to deny that trying can involve pretense of a special kind. Imagination does not function only as momentary escape from reality; I can, perhaps, imagine my way into becoming someone. Here the function of the imagination is not to fashion a substitute world, but to help us move ourselves closer to some reality we already have some grip on. I might, for instance, adopt the mannerisms of the kind of person I’m trying to be. If this were an act of aspiration, it would pain me somewhat to do so, because it is not enough for me to *act* like that person when what I want is to *be* like that person. We cannot analyze aspiration in terms of pretense because the kind of pretense we would need to invoke is an aspirational kind.

⁴³ Only possibly luckier, because there are both bad ideals and (morally) good roles for actors.

⁴⁴ I should note that not every would-be nonsmoker *aspires* to quit. It is possible to have a simpler goal of modifying one’s behavior, as in the case of Sue the overspender (see above, p.83). The aspiring nonsmoker is marked by the fact that she wants not only to behave differently, but also to come to see things differently, to cease being tempted.

6. Approximating reasons

Perhaps the value under which the pursuit is conducted is close, though not identical, to the value of the end. At the end stages of a transformative pursuit I may have access to something close enough to the final value to justify pursuit. So, for instance, I might appreciate Mozart's light operas, and this gives me reason to listen to his symphonies, and this leads me to Bach. We might try to make up a kind of series of progressively approximating values to lead the music student from music she likes to the music the class is designed to get her to appreciate. Highlights of such a series might look like this: Taylor Swift, the Beatles, Rogers and Hammerstein, Gilbert and Sullivan, Puccini, Mozart, Bach. The question is: does this series represent a subtle shift in value over time, or does it represent one single value getting progressively clarified and approximated to? Does she say, at the end, "now I see what I was after all along?"

In the first case—subtle shift—we should imagine the value transition as analogous to a move from yellow to blue along the color spectrum by imperceptibly different shades. But this is a variant of the "go for the exercise, stay for the movie" scenario. For the reason grounding the aspirant's activity when she's in the yellow region diverges from the reason in the blue region in such a way as to break up her pursuit into a series of rationally disconnected activities. From the fact that it is impossible to say where one ends and the other begins, it does not follow that there is no difference between the two. If it's a progressive clarification, there's no similar worry: the gradual shift in value would be guided throughout by the agent's sense that some target value is being approximated, like an image gradually coming into focus. But this is just what we mean in speaking of proleptic reasons. For a proleptic reason just is a reason by which an agent grasps, in an incomplete and anticipatory way, the reason that she will act on once her pursuit is successful.

Recourse to other reasons, be they approximating or vague or testimonial reasons, or reasons of pretense or self-management or competition, does not obviate the need for introducing a distinctive proleptic species of reason. I don't claim that my list exhausts all possible alternatives, but I do think it covers much of the rational territory. Moreover, there is a certain pattern that repeats itself, indicating a general strategy which the champion of proleptic reason should adopt in the face of some additional contender. If someone says, large-scale transformative pursuits can be rationalized by familiar, X-ish reasons, the proleptic reasons theorist will try to demonstrate that only a (proleptic) subspecies of X-ish reasons can hope to rationalize a distinctively aspirational pursuit.

Proleptic reasons are—I conclude—*the* reasons that rationalize large-scale transformative pursuits. A proleptic reason is an acknowledgedly immature variant of a standard reason. A proleptic reasoner is moved to ϕ by some consideration that, taken by itself, would (in her view) provide inadequate reason for ϕ -ing. But she is not moved by that consideration taken by itself; rather, she is moved by that consideration (be it competitive, testimonial, approximating, etc.) as a stand-in for another one. The proleptic reasoner uses the only valuational resources she has at her disposal, namely her current desires, attachments etc. both to mark the inadequacy of those very resources and to move herself toward a better valuational condition.

The reader may wonder why I invoke a new species of reason rather than speaking of a proleptic grasp of a (standard) reason. I do not think that much hangs on whether we attach the property of being proleptic to a reason itself, as opposed to the quality of someone's apprehension of that reason. My interest is in a set of thoughts, actions, desires, choices, and projects that neither exhibit a standard form of rationality, nor are to be discounted as irrational. The distinctiveness of proleptic rationality is my topic, whether we spell this out as a distinctive way of grasping reasons, or a grasp of a distinctive kind of reason. But there are considerations that speak in favor of the latter formulation. One context in which we might speak of proleptic reasons is to explain why someone did what he did. In this kind of case, a proleptic reason lends intelligibility to some bit of behavior. If we choose to speak of a "proleptic grasp" of a reason, then it will turn out that in proleptic cases, reasons do not explain behavior—rather, grasps do. And it is awkward to speak of actions as being explained by grasps, and natural to speak of them as being explained by reasons.

We also invoke reasons when we recommend a course of action. Suppose a mentor tells her student to ϕ in such a way as to be making a proleptic reasons statement: she can see, on the basis of what she knows about him, and her expertise in ϕ -ing, that he ought to aspire to ϕ . She cannot be read as saying that he *has* a proleptic grasp, for her point is to inform him about something he is missing. Nor is she confessing to such a grasp—for presumably, she grasps that same reason non-proleptically. We describe her as asserting that he ought to have a kind of grasp that he doesn't yet have; but that is a strange way of talking. The more natural thing to say is that she is alerting him to the presence of a special kind of reason.⁴⁵

III. A dilemma: flailing vs. deepening

⁴⁵ The distinction I am making here, between explaining an action that has already been done and recommending one as to be done, should not be confused with the distinction between motivating and normative reasons. All of the practical reasons discussed in the course of this book are (at least) motivating reasons, which is to say, reasons that potentially explain some action. (There are a number of ways of spelling out the distinction in question; here, I follow Smith 1987.) Normative reasons present some requirement on an agent's behavior, from the point of view of e.g. morality, or prudence. Many normative reasons are, of course, motivating reasons. The issue of whether there are normative reasons that are non-motivating is closely related to the question of whether there are external reasons (see below). It is worth noting, however, that all motivating reasons have some kind of claim to normative force. If I say 'S did ϕ because her neurons fired in such-and-such a way,' I offer a reason in explanation of S's ϕ -ing, but it is not a motivating reason. The reason is not S's reason for ϕ -ing; it does not rationalize her action. "A reason rationalizes an action only if it leads us to see something the agent saw, or thought he saw, in his action..." (Davidson 1980¹, p.3) Motivating reasons explain by rationalizing, and so they must present some justification, however partial, of the action. As Smith points out, motivating reasons have "the role of justifying from the perspective of the value that the very reason embodies." However, he insists on assigning to motivating reasons, "the minimal justificatory role possible," and emphasizes the possibility that the justification may only be partial: "a motivating reason, even when it does explain an agent's behavior, may reveal little of value in what the agent did even from his own point of view." (pp. 38-39) To illustrate this point, he borrows Davidson's example of a man who has always had a yen to drink a can of paint, and so he does. We have an explanation, though not a justification, of the man's action. Arguably, this example backfires: for we may not feel that we have a very good explanation of why the man in question drinks a can of paint. It seems to me that if an action's justification is only partial, the explanation, too, must be lacking in completeness. If, for instance, you ϕ -ed because you wanted to ψ , but I cannot see how anyone could want to ψ , or how anyone could want to ψ under these circumstances, or how anyone could want to ψ enough to justify the cost of ϕ -ing, then I do not fully understand why you ϕ -ed. For your action is not as intelligible to me as one without those explanatory gaps. I suspect, therefore, that motivating and normative reasons may be more difficult to separate from one another than is standardly supposed.

Let us pause to consider a dilemma for the proleptic reasons theorist⁴⁶: either the aspirant already appreciates, e.g., music to some degree or she does not. The objector posing the dilemma claims that in the first case, she is not aspiring to appreciate music but only to deepen the appreciation she already has. For if she already has a grasp of the value, then we can explain her taking the class with reference to this (ordinary, non-proleptic) reason. The objector then presents the other horn: if she does not appreciate music at all, then she cannot even aspire. She possesses too little of the grasp of the value to be, in any sense, guided by it. The best such a person could do is take the class for the wrong reason, and then be surprised to find that she enjoys it.

Let me begin with the second horn of the dilemma. I concede to the objector that without some grip on the value, aspiration threatens to devolve into what we might call flailing. Someone who simply has the idea of becoming better in some vague sense, without a grip on some specific value they will be acquiring, is unlikely to make progress⁴⁷. Consider, as an example, the wonderful chapter in *A Tale of Two Cities* in which Sidney Carton professes his love to Lucie Manette. Sidney is a self-identified drunken wastrel: Dickens describes “the cloud of caring for nothing, which overshadowed him with such a fatal darkness.” (p.261) Sidney has no intention of courting Lucie’s affections, proclaiming in one of the first exchanges of the scene that “I shall never be better than I am. I shall sink lower, and be worse.” He is grateful and relieved that she does not return his feelings:

“If it had been possible, Miss Manette, that you could have returned the love of the man you see before yourself—flung away, wasted, drunken, poor creature of misuse as you know him to be—he would have been conscious this day and hour, in spite of his happiness, that he would bring you to misery, bring you to sorrow and repentance, blight you, disgrace you, pull you down with him. I know very well that you can have no tenderness for me; I ask for none; I am even thankful that it cannot be.”

Nonetheless, Sidney takes joy in telling Lucie what she has meant to him, explaining that she awakens in him the painful glimmer of the possibility of being a better person,

“Since I knew you, I have been troubled by a remorse that I thought would never reproach me again, and have heard whispers from old voices impelling me upward, that I thought were silent forever. I have had unformed ideas of striding afresh, beginning anew, shaking off sloth and sensuality, and fighting out the abandoned fight. A dream, all a dream, that ends in nothing, and leaves the sleeper where he lay down, but I wish you to know that you inspired it.”

Lucie beseeches Sidney to aspire, telling him, “O Mr. Carton, think again! try again!” Sidney refuses, insisting that it is too late for him. Sidney can’t even get started, because he does not know quite what he would be trying to do. His grip on the good is so attenuated, so dream-like, that what he glimpses is only the bare possibility of being different. He does not actually see what value he would be acquiring or approximating to by becoming different, and so he feels that change is impossible for him. He cannot, as it were, envision himself becoming different with any concreteness. Sidney doesn’t aspire, he merely flails. It is worth noting, however, that flailing is not nothing. It means something that Sidney feels a regret he never thought he would feel again, and, although this feeling is not enough to get him to make something of his life, it

⁴⁶ I owe this dilemma to an anonymous referee.

⁴⁷ See fn. 2 for a qualification.

does move him to perform the grand gesture of sacrificing himself for Lucie at the end of the novel. Nonetheless, it doesn't give him the materials for working to become a better person. Aspiration, by contrast, really does call for some sense of the *specific* value one is guiding oneself toward. Does it, then, follow that the aspirant can simply be motivated by an ordinary reason? I do not think it does.

The objector goes wrong in assuming that the project of deepening one's grasp on a value can be explained by non-proleptic reasons. If your activity of music-listening is directed at desiring music more than you currently do, or to the extent that you currently don't, then that activity cannot be explained (non-proleptically!) by the desire you currently have. For the desire you currently have, considered in abstraction from its aspirational (proleptic) component, would only explain pursuit of the satisfaction of that desire. It cannot explain the deepening or intensification of that desire⁴⁸. The way around the dilemma, then, is that while the proleptic reasoner must indeed have some grasp on the good she pursues, her large scale transformative pursuit of it is directed precisely at the part of it she cannot yet grasp.

IV. Internal Reasons

Proleptic reasons constitute a new challenge to the thesis of *internalism* about practical reasons⁴⁹. Internalism is a thesis about what it takes for someone to have a reason to do something. Internalists hold that an agent's reasons must in some way be relativized to what she desires, where that term is construed broadly to include interests, commitments, attachments, preferences etc. First espoused by Bernard Williams (1981), internalism has since found wide acceptance, though at the same time many of those who call themselves internalists are inclined to reject some element of Williams' characterization of the position. Consider the following internalist theses:

(M) motivation condition: if R is a reason for S to ϕ , S is such as to be able to be moved by R.

(J) justification condition: if R is a reason for S to ϕ , R can be arrived at by subjecting S's set of desires to a rational procedure.

Internalists have traditionally held both (M) and (J), and expressed their combination in some formulation such as this:

(MJ) R is a reason for S to ϕ iff, were S to deliberate in a procedurally rational way from his current set of desires, he would come to be motivated to by R⁵⁰.

⁴⁸ Thus the defender of aspiration can agree with Harman (2014) against Paul 2015, that the agents described by Paul have some grip on the end-value. The claim that, e.g., pregnancy and childbirth are aspirational phenomena doesn't rest on the assumption that we are bereft of information as to what it will be like.

⁴⁹ Externalists such as Parfit (2011) go beyond negating internalism when they assert that not only are there external reasons, but *all* reasons are external, i.e., not relativized to motivation.

⁵⁰ By omitting reference to beliefs I elide the difference, here immaterial, between *subjective* and *objective* reasons. A subjective reason would be one arrived at by deliberation from the agent's current set of desires *and current set of beliefs*, whereas objective reasons would presuppose deliberating from a belief-set corrected for falsity and supplemented with any missing (and relevant) true beliefs. (see Markovits (2011b) for this way of formulating the distinction).

Internalists have wanted both to deny that someone could be in the condition of being barred from access to his own reasons, and to insist that reasons for action justify those actions in the light of the agent's desires. In short, I have whatever reasons would move me, if I were fully rational. Recently, some philosophers have called into question whether internal reasons can do both of these jobs. (MJ) lends itself to the "conditional fallacy," which amounts to a kind of blind spot for reasons that depend on one's irrationality⁵¹. Richard Johnson (1999) describes someone who has a reason to see a therapist, because he is deluded into thinking that he is James Bond. "James Bond" cannot arrive at this reason himself: for if he were in a position to reason correctly on this point, he wouldn't (so the story goes) have any need for therapy. Likewise, Michael Smith (1995) describes a sore loser so incensed by his defeat that he is inclined to punch his opponent at the end of the game. Given this inclination, he doesn't have a reason to approach his opponent at the end of the game for a handshake, though that is exactly what his fully rational, and therefore less irascible, counterpart has reason to do.

Johnson has argued that the only way around the conditional fallacy is to give up (MJ) by giving up either (M) or (J); and Julia Markovits (2011a) has recently made the case for the former option. She argues that we have independent reason to give up (M), since there are circumstances in which we aren't, and shouldn't be, motivated to ϕ by the best reason for ϕ . For instance, a pilot executing an emergency landing might be well advised not to act for the sake of saving hundreds of lives, because being motivated by this reason might put so much psychological pressure on him as to interfere with his performance of the task⁵². She advocates for a weaker version of internalism based only on (J).

I will argue that internalists—even weak internalists—are guilty of selling proleptic rationality short. But first some preliminaries. The weak internalist takes it that the reasons we have *depend rationally* on our desires. Internalists might spell out this rational dependence in a variety of ways: in terms of instrumental rationality (Hume, as understood by Williams 1981), of the presence of a sound deliberative route (Williams 1981), of the absence of rational defects (Korsgaard 1986), of procedural rationality or the reasoning of an ideally rational agent (Markovits 2011 and 2011²); of satisfying norms of consistency and coherence in such a way as to be "systematically justifiable" (Smith 1995 p.114). All of these ways of cashing out the dependence point at some analog to *formal* validity: the method in question does not add any content to one's ends, but rather takes the content already present in them and shows what reasons follow from it. The idea is: given that "James Bond" has an interest in his mental health, and also has some form of mental illness, it follows that he has reason to seek help—even if he, himself, is not in a position to appreciate this reason. Seeking help is the kind of behavior that would be consistent with the aim of mental health, when it is combined with the

⁵¹ Though Markovits (2011a) argues that one can broaden the class of counterexamples to include ones—such as Kavka's toxin puzzle, or cases where one has pragmatic reasons to hold a belief—in which the agent's inability to access the relevant reason is due not to her irrationality, but rather to certain strictures that rationality places on us.

⁵² I do have a worry here, however: in another paper (2010) Markovits argues that an action is only morally worthy if the agent is motivated by the reasons that morally justify the action. It is not clear to me how weak internalism is consistent with that view, given that Markovits presumably wants to claim that the pilot in this example is to be (morally) credited with saving all those lives. Moreover, it seems to me that the considerations she rightly adduces in favor of the conclusion of her 2010 paper—such as pointing out that we are not always aware of the considerations that motivate us—cut against those she uses, in the 2011 paper, to argue in favor of *not* being motivated by the justifying reason.

presence of mental illness. We might also speak of actions that answer or correspond to one's ends. The weak internalist might put his point thus⁵³: you have the reasons that an impartial third party observer would take you to have, if he were reasoning about what reasons you have in a procedurally rational way from your desires.

One more quick point of clarification: internalists can—and do—offer us internalist accounts both of *pro tanto* reasons and of *all-things-considered* reasons. Take Williams' (1981) example of Owen Wingrave, whose family insists that tradition gives him reason to enlist, in spite of his deep hatred of all things military. When Williams says that Owen has *no reason* to enlist, does he mean that Owen lacks even a *pro tanto* reason to do so? It is hard to imagine someone who, in Owen's circumstances, sees literally nothing speaking in favor of enlisting: surely the fact that his family strongly wants him to enlist is at least a (very weak) consideration in favor of doing so? Presumably, even if he allowed that Owen saw some (minimal) reason to enlist, Williams would still want to resist the family's insistence that enlisting is what he has an all-things-considered reason to do. For whatever glancing respect he harbors for tradition, or whatever weak desire he has to please his parents, is dwarfed by his powerful hatred of the military. In what follows, we will set *pro tanto* reasons aside: "S has a reason to ϕ ," means, henceforth, that ϕ -ing is what S has a reason to do, all things considered.

The problem is that the proleptically rational agent has a reason that not only she, but even a fully rational third party observer, will have trouble extracting from the content of his antecedent desires. Suppose the good student of music appreciation has a choice between spending an hour of her evening listening to a symphony, or devoting that hour to a hobby she thoroughly enjoys. Let us assume that listening to music will not serve any end of hers apart from her (still weak) interest in enjoying music for its own sake. The internalist must direct her to pursue the hobby she already enjoys a great deal over developing her nascent love of music. For that action coheres better with her current set of desires and interests. But if this were always good advice, we would hardly ever have reason to develop new interests, values, relationships etc.: for there is virtually always something else we could be doing that we enjoy more than, and which satisfies our other ends better than, the new form of valuation we have yet to fully acquire.

The problem is not merely that *she* does not, from where she currently stands, have a rational line of sight to the end whose value justifies her activity. For weak internalists are willing to grant that agents have more reasons than they can see their way to acknowledging. The problem is that unlike in Johnson's "James Bond" case or Smith's sore loser case, the impartial rational spectator is no better off than the agent herself. If he could somehow reason from the person's *future* condition, in which (let us suppose) love of music has become the central aesthetic pleasure of her adult life, it would be clear that she ought to listen to the symphony. But the internalist is restricted to *extracting* what the agent should do by applying a procedurally rational method to her antecedent desires, cares, interest, loves etc. The internalist must counsel us to stick with immediate and available pleasures over embarking on the arduous

⁵³ Markovits points the point in this way in a footnote (13) of her 2011² paper, though the footnote appears only in the online version of the paper, available at <https://sites.google.com/site/juliamarkovits/research>; as she points out there, both Smith (1994) §5.9 and Railton (1986) p. 174 offer re-formulations of internalism in the same vein.

process of developing a sensibility for new and perhaps higher ones. He seems to be giving us a form of advice that would have irked no one so much as Bernard Williams himself: be philistines!

My claim is that the internalist cannot capture the affective difference between the person I have called the ‘bad student,’ who is satisfied with her minimal appreciation of music, and the person who likewise harbors a minimal appreciation but aspires to become a music lover. I want now to consider some responses on the part of the internalist—some desires that he could point to in order to explain why the second has reason to listen while the first might lack it.

First, consider the desires that correspond to what I have called the reason’s “proximate face.” The aspiring music lover has promised herself chocolate for making it through the movement, and sustains her listening by imagining making a dramatic entrance in a concert hall on a snowy moonlit evening. The bad student lacks these forms of motivation. Will the internalist be able to point to these differences in their ends as accounting for the differences in their reasons? No. In order to successfully motivate oneself through some mechanism such as appetite or fantasy, the subordinate reason’s *motivational* force must outstrip that of one’s ultimate aims—but its justificatory force cannot do so. So, for instance, if I am trying to motivate myself to lose weight by promising to buy myself a nice dress, but losing weight will in fact frustrate more of my ends than it will satisfy, then my desire for a dress cannot be a source of good reasons. For the very fact that it is irrational for me to be trying to lose weight entails that it is irrational for me to be setting up incentives for myself to facilitate that project.

Alternatively, consider the class of desires that pertain, in a higher-order way, to the distal face, e.g., a desire to desire to listen to music more than one does, a desire to see what all the fuss is about, music-wise, or a desire to become a music-lover. Even if it is true that the good student has these desires, and the bad one lacks them, pointing to that difference cannot help the internalist explain the fact that the good student has a reason to listen. For the rational ground of these higher-order desires lies not in any extraneous benefit that having a stronger desire to listen to music, understanding the source of the fuss, or becoming a music lover would afford her. At least not in the case I’m imagining: someone who wants to become a music-lover in order to, e.g., please her parents, raises no problem for the internalist. For her ‘additional desire’ plugs into independent motivations that can indeed rationalize her choice in a straightforwardly internalist way. But in the case of the good student, the rational ground of her higher-order desires—the reason why she has them—is once again simply the intrinsic value of music. And this is a value she is, currently, ill-placed to appreciate. So *all* of these desires bottom out in a valuation of music that is (I posit) too weak, as it stands, to underwrite an internalistic justification of doing much of anything in its service.

Perhaps, instead of claiming that the aspirant’s reasons are based in her desires, we should allow that they might be based in her beliefs. There is a kind of internalist⁵⁴ who holds that one of the things that can rationally ground a desire (or a desire to have a desire) is a belief in the value of the object that you desire (to desire). Why couldn’t an agent’s *belief* that music is intrinsically valuable be justified independently of, and therefore underwrite, her project of changing her

⁵⁴ Namely, the kind of internalist who thinks that beliefs can give rise to desires. See Nagel (1970) ch. 5 for the canonical statement of this view.

affective response to music? If this is possible, and I think it is, then there is a version of this agent that is fully analyzable in terms of internal reasons.

The person who believes that music is valuable, but doesn't enjoy music or doesn't enjoy it very much, comes in two varieties. The first takes herself to know *perfectly well* the value of music, despite the fact that she takes less pleasure in listening to music than she thinks she could. She might work on herself to try to get herself to enjoy music more (or at all), simply for the reason that her life could contain more aesthetic pleasure than it does. Her music listening is, indeed, rationalizable by way of internal reasons—but those reasons are not proleptic, because she does not take herself to have anything to learn, value-wise. Manipulating one's affective responses so that they match the way one independently knows they should be is a real phenomenon, but it is not the one I seek to explain here.

If, on the other hand, she takes her own belief in the value of music to be in some way a defective appreciation of its value, since full appreciation would presuppose enjoyment of music, her belief will not suffice to rationally ground her attempts to access it. For she does not take her belief to already afford her (full) rational access to the value she is working to come into (better) contact with. This second case is the proleptic one that I claim internalists cannot accommodate. Such a person is willing to work harder to enjoy music than her belief can, by the logic of internalism, rationally support. Her willingness stems from her sense that there is more value out there than she has yet been able to take account of either cognitively or conatively.

Why can't the internalist simply allow that the good student has, in addition to any of the desires mentioned above, an *aspiration* to appreciate music? Internalists are famously open minded about exactly what forms of motivation or ends or conation might constitute the ground of one's reasons. I have claimed to use the word "desire" broadly—as internalists themselves often do—to cover all this whole class. They might suspect that, in this discussion, I have actually used it more narrowly, in such a way as to unfairly exclude the one kind of pro-attitude relevant to differentiating the good student from the bad one. But this is not the case. I do not want to deny the internalist recourse to the concept of being disposed to be motivated in a way that outstrips the reasons derivable from their current motivational set. The problem is that she cannot make room for the fact that any of those motivations are *rational*. For the internalist, letting 'aspiration' into one's subjective motivational set simply means letting in a tendency to be motivated in an incoherent and procedurally irrational way. What the internalist cannot do is to derive the good music student's reasons not merely from her aspiration but her *rational aspiration*. For her theory, as I've been arguing, gives us no way to see how that phrase could be anything but an oxymoron.

At this point, we may feel some nostalgia for old-school internalism. Markovits ascribes reasons to me on the basis of the what a third party, impartial, perfect reasoner would take as answering to my present motivational condition. Williams, by contrast, is interested in what reasons I, with all my imperfections, could arrive at. It is true that Williams must understand what I "could arrive at" in a way that includes the concept of rationality, i.e., as "could *rationally* arrive at"—but he nonetheless has a broader and in a certain way softer construal of what it means to arrive rationally at some conclusion. He doesn't seem interested in specifying a procedure that could be vouchsafed as formally valid, and therefore employed an identical form

by *any* rational agent. Rather, he seems to want to claim that an agent must be in a position to somehow or other see her way to any reason we are to count as her own. Hence his famously—to some, aggravatingly—open-minded conception of what such ‘deliberation’ consists in: “practical reasoning is a heuristic process, and an imaginative one, and there are no fixed boundaries on the continuum from rational thought to inspiration and conversion.” (1981, p.110)

Williams’ followers have tended to be much more restrictive than he was in what they are willing to count as rational deliberation. It has seemed to some that without such restrictions it is not clear what the theory means to rule out, and thus what the contrast with externalism is meant to amount to. Others have harbored substantive worries about some of the forms of reasoning that Williams wants to admit. For instance, Smith objects that “the imagination is liable to all sorts of distorting influences, influences that it is the role of systematic reasoning to sort out.” (1995, p.116)⁵⁵. Finally, as I observed above, the conditional fallacy has driven still others (e.g., Markovits) to place at the heart of internalism the idea of what can be deduced by a valid procedure from a given set of desires.

Whatever the disadvantages of Williams’ internalism, it might seem to be in a better position to accommodate proleptic reasoning than weak internalism. Indeed, I believe Williams himself may have thought that by emphasizing the role of the imagination in reasoning, he was skirting the worry about philistinism I’ve been pressing here. When Williams warns against an overly narrow conception of what a ‘sound deliberative route’ may consist in, reminding us that “the imagination can create new possibilities and new desires,” (1981, pp.104-5), he may have large scale transformative pursuits in mind. For it is true that we use our imaginations to grasp the value that a radically new form of life has to offer us. The problem is that we cannot do so well enough to generate an internal reason. The music student uses her imagination to generate a fantasy about a snowy evening, and this imaginative work may well be crucial to her forward progress. But she cannot, in fantasizing in that way, foresee the *real* value that music will bring for her. Imagination simply doesn’t have that power. No matter how loosely we hold the reins, deliberation will not plot a course from the agent’s present condition to what I have called the distal face of her proleptic reason. We cannot attribute to the aspiring X-er imaginative or heuristic resources that so far outstrip her current motivational condition that she is able to imagine her way into the intrinsic value of X.

Internalists may respond to this line of reasoning by beginning to doubt whether they *want* to accommodate proleptic rationality. There is no knowing whether an agent’s course of action *will* end in ϕ -ing until the course has, in fact, ended. Are we to ascribe proleptic reasons only retrospectively, on the basis of successful ϕ -ing? Internalists may raise the same kind of objection to recognizing proleptic rationality that Smith raises to Williams’ idea of the imagination as a source of reasons. They may doubt whether there is a fact of the matter as to whether what an agent does in the service of such an indeterminate goal is, or is not, proleptically rational. They may question whether it is even possible to ascertain that someone who takes herself to have a proleptic reason in fact does not, or vice versa.

⁵⁵ This for a variety of reasons.

I grant that the early stages of value-acquisition may indeed be tenuous enough to be immune to rational critique. Aspiration begins as something like wish or hope, and we would tend not to tell someone she ‘shouldn’t’ have such and such a long-term wish, or that her cherished hopes for her future self are ‘irrational.’ Rational criticism does, however, eventually become appropriate. At some point on the way to her goal, the agent enters a space in which it becomes fitting for someone—though perhaps, not just anyone—to say either ‘try harder, you can do this’ or ‘give up, this isn’t working for you.’ These are the kinds of locutions by which we key someone in to the presence or absence of proleptic reasons. We can see the direction someone is heading, assessing her trajectory on the basis of the work she has done so far. We gauge whether she has it in her to make it to the endpoint, whether it is reasonable for her to proceed, or more reasonable for her to try something else. Or, rather, those of us with the relevant expertise and the relevant familiarity with the aspirant do this.

Though proleptic reasons are amenable to rational critique, the character who is in a position to offer this critique is not Markovits’ impartial, detached, perfectly rational observer. This observation may further incline the internalist to reject the rationality of proleptic reasons, but I think it should instead lead her to question the unargued-for assumption that the “perfectly rational agent” is the perfect arbiter of all practical reasons. If it were true that excellence with respect to procedural rationality alone—a kind of analytical prowess—put someone in a position to determine what reasons a person has, philosophers would be much better at offering advice on any sort of practical topic than we in fact are. It is important to keep in mind that the set of examples with which we philosophers discuss practical rationality does not represent a random sample. Philosophers tend, quite reasonably, to gravitate toward examples that provide immediate spectatorial access. The “impartial rational observer” can determine without wanting anything, doing anything, or having any special expertise, that breaking an egg is a rational means to the end of making an omelet, and that leaving the egg intact is not a rational means to the same end. In order to make the relevant determination, all one needs is an understanding of what eggs are, and what omelets are. When speaking to an audience—philosophers—without any special practical competence, it is useful to avail oneself of examples that *can* be assessed by any rational observer.

But we should guard against taking such armchair assessability to be a feature of practical rationality itself. For instance, consider the difficulty of determining whether it is an intensive course, years of casual listening, or a season of concert-attendance that represent the rational means, for the would-be music lover, to realize her aspirations. One doesn’t know the answer to this question merely by knowing what the relevant items *are*. And not even a master of procedural rationality should, I think, venture to answer this question if she has never had any interest in music.

At least some forms of practical rationality or irrationality may only be evident to those whose sensibilities—desires, emotions, intellects—have been shaped by the practice in question. In addition, such judgments often call for personal acquaintance with the subject whose proleptic rationality is being called into question. And even when an expert is assessing a subject she knows well, she will often be unable to judge whether the aspiration is rational or not until she has some actual extent of practice before her. Thinking about whether or not something will work out is not always a reasonable substitute for trying to work it out. It does not tell against

the rationality of aspiration that a judgment as to whether someone has a proleptic reason is likely to be made on the basis of something like a trial period, or evidence of similar past attempts, and that it is likely to call for personal acquaintance with and personal affection for both the subject in question and her aspirational target. Judgments of practical (ir)rationality sometimes call for practical experience.

We acquire most, perhaps all, of our practical knowledge by responding to past experience. My interest has been in those cases in which the experience that we respond to is one that we ourselves have sought out; moreover, we sought it out for the (proleptic) reason that it produce this response. In those cases, we have guided ourselves to the new values or desires or commitments that our experience engenders. That process of self-guidance is a kind of practical learning. Because a process of learning some new form of valuation is not the same as a process of articulating or rendering consistent the values one already has, proleptic reasons break every internalist's mold.

*Chapter 3: Intrinsic and Extrinsic Conflicts of Desire

In what I will call an “extrinsic” conflict of desire, an agent’s desires pull her toward incompatible actions. Such an agent’s problem is that, as a matter of contingent fact, nothing she does will get her everything she wants. There is, however, another kind of desire-conflict, in which the agent’s desires pull directly against one other. For instance, generously hoping for someone’s happiness gets in the way of resentfully wishing to see him suffer. In a loving-but-spiteful moment, I might be torn between those desires. Harry Frankfurt called our attention to special difficulties attending the resolution of this kind of conflict, arguing that we do so by *identifying with* the one desire and *externalizing* the other. In this chapter, will show that Frankfurtian identification/externalization cannot resolve such conflicts, which I call “intrinsic,” and propose an alternative: they are resolved by aspiration. I begin with a brief overview of the argument of this chapter.

Frankfurt does not offer an account of what makes a conflict intrinsic. Taking my bearings from his characterization of the examples he offers of the two sorts of conflict, I argue that in order to get the conflict properly into view, we must, first, grasp that the desires of the intrinsically conflicted agent form parts of intrinsically conflicting *values*. A conflict at the level of value fractures the agent’s evaluative perspective: in order to get the desirability of one of the things she wants fully in view, she must step out of the point of view from which the other appears desirable. I concede that, on this analysis, it becomes difficult to see how intrinsic conflict is even possible. How can an agent simultaneously inhabit incompatible points of view? I draw on a parallel with cases of split attention: the desires cut against each other, in that she can have the one desire to the degree that she doesn’t have the other. One can experience both of two intrinsically conflicting desires, insofar as one experiences each of them in a qualified way.

Though I ultimately deny Frankfurt’s positive conclusion that intrinsic conflicts are resolvable by identification, I offer an argument for a related negative conclusion: intrinsic conflicts are *not* resolvable in the way in which we standardly resolve extrinsic conflicts, namely by deliberation. By articulating this argument, I put Frankfurt into conversation with a number of ethical theorists who have aimed to show that there are value-conflicts deliberation cannot resolve. Arguments establishing the limited power of deliberation have been offered in support of Aristotelianism (by John McDowell), and against both Kantianism (by Bernard Williams), and Consequentialism (by David Sobel). What is distinctive about Frankfurt’s contribution to this series of arguments is that he is committed to the idea that desires whose conflict is deliberatively irresolvable are nonetheless psychologically compossible. Thus it is for Frankfurt, and not for the others, that the question of the *resolution* of deliberatively irresolvable conflict emerges.

I distinguish two ways in which a conflict might not be resolvable by deliberation: it might be, first, that the agent cannot use deliberation to *answer* the question as to which of two options she ought to choose; or alternatively, it might be that she cannot even *ask* this question. Frankfurt’s solution to intrinsic conflict—identification—is predicated on the mistaken assumption that intrinsic conflicts are deliberatively irresolvable for the first reason. For he takes it that intrinsic conflict comes to an end when the agent surveys her two desires and

decides between them. In fact, I argue, an agent who is intrinsically conflicted cannot achieve the reflective distance from her conflict that is necessary in order to even arrive at the state of uncertainty that would have to prefigure such a decision. Can such an agent resolve her conflict at all? I argue that she can do so by aspiring. Aspiration, the temporally extended process by which someone works to become a different and better person, is the solution to intrinsic conflict.

I. Framing Intrinsic Conflict

Harry Frankfurt (1976, p.248, cf. 1988 p.170) differentiates what he calls “two sorts of conflict of desire.” He illustrates the distinction with a pair of agents roughly along the lines of these:

TORN AESTHETE: A man would enjoy both attending a classical concert and going to the film that is playing at the same time as the concert. Frankfurt says that he “would resolve the problem this conflict presents just by deciding which of the two things in question he prefers to do.” If he decides on the film, but cannot get a ticket, “it would be quite natural for him to revert to his second choice and go to the concert.” (1976, p.249)

BITTER WIFE: A woman wants to mail her husband’s letter, as a favor to him. But they have a complicated relationship. She loves him, and she knows it will make him a lot happier if the letter is mailed—unmailed letters are one of his pet peeves. But she is also intensely bitter and angry at him for what she perceives as his many small cruelties toward her, his intense irritability (so many pet peeves!), his lack of romantic initiative. Next to the mailbox she sees a garbage can, and it occurs to her to spitefully throw his letter in the garbage instead of mailing it⁵⁶. Suppose that, in the end, love prevails over spite, and she reaches for the handle of the mailbox—but finds it locked. Frankfurt notes that “This would not naturally lead h[er] to see if [s]he could salvage the satisfaction of h[er] other desire.... the alternative of injuring [her husband] is not second to the person’s first choice of [doing him a favor].” (1976, p.249)

Let us call the aesthete’s conflict “extrinsic,” and the wife’s “intrinsic.” As a first pass at explicating this distinction, we might suppose that the difference lies in whether the intentional contents of the desires are themselves opposed. The wife, in desiring to hurt her husband, desires the opposite of what she desires when she desires to help him. In desiring to hurt him, she desires that he fare badly, whereas in desiring to help him, she desires that he fare well, i.e., *not* badly. The aesthete, on the other hand, doesn’t desire *not* to see a movie. He desires two things—to see a movie, and to see a concert—whose conflict is only ‘in the world.’ It is not wrong to say that intrinsic conflicts feature such internal contradictions, but offering this as the basis of the distinction is too quick. For there is a problem about how to individuate the desires that show up in these conflicts.

It is possible to pick out the intentional content of a given desire anywhere along a continuum of objects that ranges from certain maximally abstractly formulations (“happiness,” “acting well”), on the one hand, to a particular action, on the other. We can exploit this fact to create the appearance of contradiction amongst extrinsically conflicting desires, or the appearance of a

⁵⁶ I’ve modified Frankfurt’s original example, in which the spiteful act is a speech act, and it is directed against a casual acquaintance. I cite the original example below, p. 74.

lack of contradiction amongst intrinsically conflicting ones. Someone who desires, e.g., both that the table be heavy and that it not be heavy, has what are (at least one some sense) contradictory desires. But perhaps he desires it to be heavy because he does not want it to easily tip over, but he desires it to be light because it wants it to be easily moveable when he relocates. His conflict is extrinsic, despite the fact that we can describe him as someone who desires that contradictory propositions be the case. Redescribed, what he wants is something that is both light and moveable, and there is no contradiction there. Consider the aesthete again: insofar as he resolves to go to the movie, and he sees that this entails not attending the concert, we may say he desires not to attend the concert. His desire to attend the movie *is* a desire not to attend the concert. In the case of the wife, we could ‘hide’ the intrinsic quality of her conflict by describing it as a desire to mail the letter vs. a desire throw it in the trash.

In order to get the intrinsic quality of the wife’s conflict into view, we must also rule out certain kinds of backstories explaining her desires. Suppose, for instance, that she wants to throw the letter away to teach him a lesson about taking her for granted; it is part of a (perhaps ill-conceived) plan to save their marriage. She is hurting him in order to help him. Alternatively, she knows the letter is full of damaging revelations—she is ‘helping’ him in order to destroy his reputation. If she is helping him in order to hurt him, or vice versa, then the options of mailing and trashing are, as for the aesthete, preferentially ordered. Should it turn out that she cannot hurt (help) him by mailing (trashing) the letter, and she can do so, to a lesser degree, by trashing (mailing) the letter, it would be quite natural for her, in Frankfurt’s phrase, “to revert to her second choice.” Likewise if both desires serve some third value, such as the promoting of the wife’s political career: “On the one hand, male voters like him. On the other hand, female voters hate him. And they’re the ones I really need to court at the moment. If I work things so that he initiates the divorce, I’ll get a sympathy bump as well.” It is true that she wants to hurt him (in order to court female voters) and to help him (in order to court male ones), but that is not really the level at which her conflict lies. Here again, it is easy to imagine her falling back on her second best option: if she cannot succeed with the female voters by trashing the letter, she will court the men by mailing it.

Specifying intrinsic conflict requires more than pointing out that an agent has a desire to ϕ and a desire not to ϕ . We need to add that this level of description picks out the level at which she is really conflicted. In the next section, I propose that we can do this by shifting ground: instead of considering the two desires in isolation, we must situate them in the broader evaluative outlooks of which they are a part. The problem for the wife is that her two desires conflict at the level of value.

II. Valuing: a hybrid account

Like Niko Kolodny (2003), Samuel Scheffler (2011) and Jay Wallace (2013), I deny that valuing can be simply identified with believing or with desiring. I share their hybrid account, on which valuing includes both cognitive and conative elements. Scheffler explains (p.21) why it is a mistake to think that we value everything we believe valuable: our capacity to believe that things are valuable far outstrips our capacity for personally investing ourselves in those objects we can truly be said to value. In order to value something, we must engage with it in a way that takes time, effort, and practice. A creature with a finite lifespan and resources cannot devote herself to all of the things she sees as valuable.

Wallace, drawing on Scheffler, has emphasized the ways such devotion manifests itself in affective connections between the subject and the valued object. He describes the valuer as “subject to a range of characteristic emotional reactions, depending on how things are going with the object of concern.” (p.23) Coming to value something entails opening yourself up to being hurt in relation to that thing: you become vulnerable to forces that threaten the valued object, or threaten to separate you from it. But valuing involves more than caring: when we value something, we also *evaluate* it as in some way good or worth caring about. It is possible to be emotionally vulnerable to something one thinks badly of, and wishes one did not care about. When I value something, by contrast, I approve of my own affective entanglement with it. Thus valuing seems to include *both* an affective component and a cognitive one. But there is more.

Kolodny (p.150) has drawn attention to the connection between valuing and the recognition of practical reasons. When one values something, one is motivationally disposed in relation to that thing: to protect it, preserve it, engage with or in it. Someone who values something is not only affectively responsive to it but also motivationally engaged with it; the things I value delimit the shape of my practical rationality, in that my values determine what shows up for me as a reasonable thing to do. For instance, if I value philosophy, then I am disposed to respond motivationally to opportunities to engage in it. Likewise, if I value my relationship with my friend, I am disposed to, e.g., respond to her emails, call her, make plans to get together. If I believe philosophy or my friendship to be valuable but am not inclined to engage in philosophy or make any effort to be in contact with my friend, then I do not really value either one. And this is true even if I have an emotional vulnerability that manifests in, e.g., feeling guilty about the fact that I never do any philosophy or call my friend.

Scheffler notes a fourth element to the account of value. In addition to the cognitive, affective and motivational ways in which we respond to objects of value, Scheffler (p.29) says that valuing includes an element of self-monitoring: when we value something we react to our own responses to the valued object, experiencing the affective, conative and motivational responses described above as merited or appropriate.⁵⁷ In order to show that this is indeed an additional condition, not entailed by the first three, let me offer a few examples.

(1) I find myself developing a passion for motorcycle riding at a period in my life when I cannot afford the investment of time or money such an activity demands. In addition, my children are young, and I find it irresponsible to risk my safety. I genuinely believe that riding a motorcycle is a valuable activity, and I don't think I'm wrong to be thrilled by the prospect of doing so, but I experience my own motivational disposition to engage in the activity as inappropriate: now is not the time.

(2) An older brother finds himself with a growing passion for magic just as his younger brother is displaying talent in magic. The older brother might judge that magic is valuable, be motivated to do it, feel excited at the prospect of pursuing it and saddened at the thought of giving it up. Nonetheless, he might judge that these reactions are not warranted. He might

⁵⁷ In fact Scheffler makes a somewhat more restricted claim: that we experience the emotional or affective response as appropriate. As my examples show, however, one can extend this condition to cover motivational and cognitive responses as well.

recognize that while magic really is valuable, his own passion it is suspect, a manifestation of his reluctance to allow his brother an arena in which to shine.

(3) A photographer, looking at a photograph of a gruesome scene, is struck by its beauty. She takes pleasure in the beauty of the image and is initially inclined to put it up on her wall when she checks herself: “What is wrong with me?! I ought not put this up on the wall, I ought not take pleasure in it, and I ought not even judge it to be beautiful.”⁵⁸

In the first case, the agent disapproves only of the motivational response, in the second of the motivational and affective responses, and in the third, of all three. These are all, plausibly, cases in which we want to deny that the agent is engaged in the valuing activity in question. When an agent *does* engage in valuing, by contrast, she approves of her affective entanglement with the object. These responses indicate the presence of a self-monitoring component to the activity of valuing: the valuer makes sure that her various responses to the valued object fall in line. The valuer thus experiences her own motivation to pursue, protect and engage with it as fitting; she will not, as in my third example, feel alienated from her own assessment of the object as valuable. The self monitoring activity also has a negative side. Insofar as she is disposed to experience the positive responses of pursuing, protecting, enjoying etc. as appropriate, that same disposition will bring her to reject, disapprove of and feel alienated from a *contrary* set of attitudes: feelings of disgust or hatred or indifference toward the object; motivations to destroy, avoid or disengage from the object; beliefs that the object is worthless, evil or in some other way not-valuable. And this brings us to the bitter wife.

Her desire to mail the letter is not a stray impulse, but part and parcel of a larger complex of valuing. She has feelings of warmth and affection for her husband, is motivated to do things that will make him happy, and believes in the value of their relationship. And she experiences all of these reactions as appropriate responses to the value of their relationship. She thus satisfies the conditions on valuing her relationship with her husband⁵⁹. But this does not prevent her from *also* experiencing reactions of spite or hostility toward her husband: she is motivated to do things to damage their relationship, she has feelings of contempt, she believes that she is being exploited and that her husband is undeserving of her love. This set of responses sits ill with her valuation of her relationship: insofar as she values her relationship, she cannot but experience them as inappropriate. We cannot, however, confuse this fact with the idea that she *doesn't* value her husband or their relationship.

An intrinsically conflicted agent experiences a desire (or feeling, or belief) that her values bring her to see as inappropriate. The stray desire throws a wrench in the works of her valuational machinery. We can now compare the wife to some of the non-intrinsically conflicted variants I sketched above. If the wife were inclined to throw the letter away in order to teach him a lesson, as part of a plan to save her marriage, she would not be intrinsically conflicted because she would not experience that motive as inappropriate to her valuation of their relationship. If, by contrast, she sees both options in terms of her political career, then she would not be intrinsically conflicted because she does not value their relationship in the first place. In this example, she values her political career, and this does not require her to reject either of her two

⁵⁸ This example is inspired by the GERMAN ARTIST & 9/11

⁵⁹ For the purpose of analyzing this example, I adopt Kolodny's construal of love as the valuation of a relationship. I should note that nothing in my argument hangs on this particular analysis of love.

impulses. Finally, the aesthete is not intrinsically conflicted because his valuation of music does not require him to disapprove of his desire to see a movie; nor does his appreciation of cinematic value call for a rejection of his desire to attend the concert. The aesthete's conflict consists in the fact that *pursuing* or realizing one of his values gets in the way of his *pursuing* or realizing his other value. His problem is that he cannot *do* or *achieve* or *bring about* both. The wife's problem is not, in the first instance, that mailing the letter gets in the way of trashing it, or vice versa—though this is of course true—but that she has difficulty even *wanting* both of the things she finds herself wanting.

The intrinsic character of the wife's conflict emerges when we situate her desires in the larger framework of her values. If two attitudes constitute an intrinsic conflict, this is because one of them is part of a larger unit of the agent's psychology—a value—where that larger unit, in turn, calls for her to regard the other attitude as inappropriate. The evaluative nexus of love calls for her not to desire to trash the letter. Her hatred, too, seems to represent a form of valuing: she has the relevant evaluative beliefs (her husband is unromantic, thoughtless), affects (resentment, spite) and motivations (to trash the letter); moreover, in her spiteful moments, these strike her as exactly the appropriate reactions to him, and their relationship. Her hatred is, thus, an organized evaluative nexus in just the sense that her love is; and from the point of view of that hatred, her desire to mail the letter makes little sense.

Torn between loving attention that manifests her valuation of her relationship, and spiteful revenge that inclines her to destroy it, the wife suffers from a kind of double vision. She sees the goodness of mailing with one eye and the goodness of trashing with the other. But she cannot get both of these 'goods' into view at once. There is no problem, for the aesthete, in admitting that cinematic and musical value are both forms of value. The wife's choice, by contrast, cannot but fracture her evaluative point of view. For her motivation to mail the letter, given the valuational network in which it is integrated, puts a kind of pressure on her *not* to be motivated to trash it. Her conflict seems to be situated at the level of the values themselves, because her conflict is between items she cannot value together. Two desires conflict intrinsically if their conflict divides the agent's evaluative point of view against itself. Experiencing these two desires seems to require that she be two valuers at once.

III. The Possibility of Intrinsic Conflict

The wife cannot, of course, actually be two valuers at once. How, then, is intrinsic conflict possible? The wife is poised between thinking of the prospect of mailing her husband's letter in a loving, forgiving, generous way, and thinking about it in a bitter, spiteful, vengeful way. Someone who takes pleasure in her husband's comfort seems to be a different kind of person from someone who takes pleasure in his suffering. Could someone *feel* the generous desire to make a person happy and at the same time equally and fully *feel* the spiteful desire to hurt that person? Having one of these feelings seems to get in the way of having the other. For that is precisely what we mean by saying that her conflict is intrinsic: her valuation of her relationship militates not only against satisfying, but even against feeling, her spiteful desires.

And yet we should not concede the impossibility of synchronic intrinsic conflict; for we will find that we cannot easily redescribe such conflicts in terms of diachronic vacillation. Consider

what happens if we try to compose the bitter wife out of alternating periods of loving-wife and spiteful-wife. We will find ourselves imagining someone who wholeheartedly loves and cherishes her husband at one moment, then switches to spitefully plotting his demise, and then flips back to adoring love. That is a person suffering some deep psychological pathology. Spite and love in a mentally healthy person inevitably color each other. The fact that the wife also wants to hurt her husband cannot but have an impact on, i.e. detract from, the quality of her love for him. I do not want to deny that the wife might go back and forth between the mailbox and the trash can, driven at one moment by love, at another by spite. But the vacillation that serves as a familiar marker of intrinsic conflict is not of the Manichean kind to which an eliminative reduction of synchronic intrinsic conflict would have to restrict itself.

Recall what it is like to fight with someone you love. It would not be surprising to hear anger in her voice. It would be shocking to hear anger *unaccompanied by love*. The latter is the experience of being the object of contempt or loathing; and that is not the experience the husband would have, were the wife to express her feelings to him as she heads toward the trash can. Likewise, the love that comes to dominate her thinking as she turns back to the mailbox is love of a particular kind: the love of someone who struggles against hate, who loves by trying to love. The wife's conflict, even when it involves vacillation, presents in the form of love marred by spite, or spite inflected by love. We must, then, examine how the value condition of love be mixed, synchronically, with that of spite.

Let's start by considering ordinary cases of divided attention. Suppose that I am trying to work through the argument of a paper while my child cries loudly in the background. One thing that might happen in this situation is that the presence of the one prevents me from attending to the other altogether. I might be so absorbed in the paper that I don't register my child crying, or so distressed by my child's crying I forget about the paper. But I also might be in a condition between these two, a condition of split attention. I know my partner is handling the baby, so I set out to focus on the paper, but my grasp of the argument is hazier than it would be, were the room silent. Or I am holding the baby, and attending to her needs, but not as well as I would be, were my mind not on the paper. In these cases, the quality of my attention to the one suffers in virtue of my attention to the other. Note that attention does not always suffer from being divided: I might be perfectly capable of cooking a dish I know well while thinking about the paper, but incapable of attending to my child while thinking about the paper. In cases of the second kind, we are moved to introduce 'degrees' of attention or awareness. And this, I suggest, is how we should understand the case of the wife.

A purely loving wife thinks about how to make her husband happy; a purely spiteful wife takes undiluted pleasure in the prospect of his pain. The bitter wife's thoughts incline in both directions at once. She cannot fully devote her attention to either value—the value of his happiness, or that of his misery—because she is always distracted by a demand to look at the world in an incompatible way. The struggle she experiences in virtue of her mixed condition is manifest in every aspect of her agency: her facial expressions, her idle movements, her inner monologue, her feelings, her speech.

People in situations such as the wife's exhibit some of the behavior we see in cases of split perceptual or cognitive attention⁶⁰. They aim to 'silence' one desire and raise the volume on the other. So the wife might, in a more loving moment, avert her eyes from the trash can, or make a mental note to stop associating with recent divorcees. When memories of his behavior that morning begin to ignite the familiar flame of indignation, she takes a deep breath and thinks instead about something nice he did for her. In a spiteful moment, she does the opposite: calling up friends she knows will supportively encourage her to vent her bitterness, replaying the details of their morning fight in her mind. We can imagine parallels with the baby/work example: putting on noise-cancelling headphones so that I can attend to the paper; or, alternatively, resolving not to bring work home so that I can attend to the baby.

My proposal is that we can make room for having two conflicting value perspectives at the same time by admitting the possibility of possessing each value in some degree. Noticing a jealous impulse in oneself can't but diminish the quality of one's generous desire to help; acknowledgement that one does still love one's husband cannot but get in the way of fully reveling in spiteful pleasure. If we allow that each value is incompletely or imperfectly present to just the degree that the other is present as well, conflict can be the object of experience. Intrinsic conflict is possible if the conflicting evaluative perspectives are psychologically compossible; and they are, but in a qualified way.

IV. Deliberative Irresolvability

Frankfurt argues that intrinsic conflicts are resolved by *identification* with one of the desires and *externalization* of the other(s). He does not specify the point of contrast with the resolution of extrinsic desire, but the natural one to supply is that extrinsic conflicts are resolved by deliberation: I decide which of the two desires should, given relevant contextual considerations, be favored over the other. Frankfurt hints at such an answer when he notes that desires which conflict intrinsically "do not belong to the same ordering" whereas, in the extrinsic case, it "would be quite natural for him to revert to his second choice." I have articulated a conception of intrinsic conflict on which it is a conflict between two valuational outlooks that cannot constitute a single point of view. I want now to argue that this way of conceiving of intrinsic conflict supports Frankfurt's intuition that intrinsic conflicts call for a non-deliberative mode of resolution. But it is important to say more than this: for there are deliberatively irresolvable conflicts that are not intrinsic. I want to isolate the particular species of deliberative irresolvability specific to intrinsic conflicts.

⁶⁰ I cannot offer a precise account of the relation between the nature of the attention split in desire-conflict cases, on the one hand, and that in dual-task cases. My thought here is only that looking to the latter can shed light on the more general phenomenon of qualified psychological compossibility. It may be that the desire-conflict case only shares some structural features of the dual-task cases, but deep differences render the use of the term 'attention' for desire-conflict, if not homonymous, then at least metaphorical. Or, alternatively, it may be that there is one genus of attention of which both of these are species. The second point would have to be established in the light of a general analysis of the practical import of attention (on which, see Wu 2011). Since I cannot undertake such an analysis here, all of my references to ethical "attention" should be understood in the first, more cautious way.

When we deliberate, we ask which of two (or more) things we ought to do⁶¹. This is a comparative inquiry, and the comparison in question is one of value. Deliberation asks us to compare the (anticipated) value of one option with that of another. The wife cannot ask which is “best overall” because there is no single conception of goodness that will accommodate her two sets of values. When she thinks about the ways and degrees to which causing her husband pain makes her happy, she is conceiving of her own happiness in a radically different way from the way in which she conceives of it when she thinks about how doing him a favor makes her happy. The point of view from which making her husband happy is an appealing prospect—a sincere, loving, desire for his well being—is one from which making him miserable seems in no way good. And vice versa. The two options cannot be surveyed because accepting the deliberative relevance of the one entails denying that of the other. She cannot deliberate from both points of view at once.

Consider, by analogy, the familiar form of speech act that runs like this: “As your friend, I say go ahead, but as your attorney, I advise against it.” Suppose that what I mean by this is that I cannot tell you what to do, all things considered. When I look at your situation from the point of view of friendship, a different set of values show up from the ones I see when I look at it in my capacity as an attorney. No third perspective is available to me, with reference to which I could impartially assess both sets of values. I can only offer you the guidance of helping you weigh the considerations in each value-space. I can tell you, e.g., how the legal reasons add up. The last step—that of weighing or otherwise adjudicating the two sets of values—I must leave to you.

Being your friend is an advisory role, which is to say that it puts me in a position to be of deliberative assistance to you; being your attorney, likewise, constitutes such an advisory role. In the case I’m imagining, what I’m telling you is that being your friend-and-lawyer is not an advisory role. It does not represent a position from which I might make recommendations to you. In such cases, we often worry that the non-professional perspective will ‘infect’ the professional one and make it difficult for the person to give good professional advice. It is sometimes wise to avoid getting medical or legal advice from close family or friends, exactly for this reason: such situations present the advisor with what we call a “conflict of interest.” Conflicts of interest are interpersonal analogs of intrinsic conflicts.

In the case of intrinsic conflict, we could say that the agent experiences a conflict of interest between her role as loving wife and her role as hateful wife. She is in a situation in which the deliberative question, “which of these two thing should I do?” cannot be raised. We only ask ourselves whether we should do something when that thing seems good to do. And in this case,

⁶¹ I do not want to deny that practical reasoning may, at times, resolve the question of *how* to bring some proposed goal about in a thoroughly noncomparative way. (See Nielsen 2011 for an argument that Aristotelian *bouleusis* was primarily noncomparative.) We need not always, at every turn, insist on doing what we have ascertained to be better than all other available alternatives. I may, for instance, reason as follows: “How will I get to Paris? By taking a plane. Where do I get a ticket? Online.” I can, thereby, reason to the action of turning on my computer without, e.g., having considered and rejected the option of consulting a travel agent. Straightforward implementation of a pre-selected plan is, no doubt, one function of practical rationality. But it is not a function relevant to the topic of intrinsic or extrinsic conflict of desire. For that topic calls for a form of rationality that might guide agents torn between a multiplicity of available options. Hence my focus on *comparative* practical reasoning, for which I reserve the word deliberation here.

the one option only seems like a good answer to the question of what to do insofar as one is deliberating from a point of view on which the other option doesn't. We can see the distinctiveness of this kind of quandary by comparing it to one in which the deliberative question is difficult or even impossible to *answer*. As Donald Davidson says, "The situation is common; life is crowded with examples: I ought to do it because it will save a life. I ought not because it will be a lie; if I do it, I will break my word to Lavina, if I don't, I will break my word to Lolita; and so on." (1980, p.34) Let us add to Davidson's list the familiar examples of Sophie's choice between the lives of her two children, and Sartre's resistance fighter. (He can either fulfill his duty to take care of his aging mother, *or* his duty to fight for his country.)

In these cases—call them 'dilemmas'—the agent experiences unhappiness no matter which choice she makes. Furthermore, she may feel she has no non-arbitrary means to decide between her options. And this might be not a contingent but a necessary fact about the values in question—for we might suppose that the values of mother and motherland (or fulfillment of the two promises, or the lives of the two children, or the movie and the concert) are incommensurable⁶². In that case, there is no fact of the matter as to which ought to be preferred by the agent who must choose between them. An agent placed in a dilemma between incommensurable values is in a difficult situation, no doubt, but her conflict is not intrinsic. It is no part of loving one child that it is inappropriate to love another; nor does one's valuation of one's country make it unacceptable to value one's mother. When one considers what one should do, both options (e.g. saving mother and saving country, or saving the younger and saving the older child, or keeping the promise to Lavina and keeping the promise to Lolita) *show up* as valuable things to do. The problem lies in figuring out which is more valuable. If there is value incommensurability, this question might be unanswerable; but that is not the same thing as saying that the question cannot be asked.

The problem of value incommensurability, if it is one, is a problem for a subset of *extrinsic* conflicts: ones in which the (purportedly) incommensurable values do not stand to one another in any of the canonical value relations (see Chang 1997). Neither one is better than, worse than, or equal to the other. Being in this situation, and recognizing that one is in it, may make it impossible to come up with a fully satisfying answer, or any answer at all, to the question "which of the two should I pursue?" Note that the difficulties in answering this question are not restricted to cases of incommensurable values. Some choices are beyond my individual capacities for commensuration, or would take more time than I have available to me for making

⁶² I follow conventional usage in treating 'incommensurable' as a synonym for 'incomparable,' where both are to be understood as having the meaning, 'cannot be compared.' This contrast's with Chang's (1997) usage, since she restricts the term 'incommensurable' to describing a species of incomparability. (Commensuration, in Chang's terminology, concerns the question of whether we can say *how much* more of a given value the one item has than the other.) Nothing I say here is in substantive conflict with Chang's treatment of either the genus or the species; my contention that there are deliberatively irresolvable conflicts is not meant to call into question her denial of (what she calls) incomparability. For Chang argues that incomparability and comparability are three place relations, relating one item to another in terms of a third—which she calls the 'covering value.' The covering value is the value in terms of which the two items are to be (or cannot be) compared. If, as in my cases of intrinsic conflict, no covering value is nominated, then Chang holds that it is improper to speak of the items as being (in my terms) commensurable or incommensurable. The most we could say is that the items in question have not been compared.

the decision. Or I may successfully commensurate the two options as being exactly equal⁶³, in which case I still lack an answer to the question of what to do⁶⁴. In none of these cases—including that of incommensurable value—does it follow from the fact that I cannot deliberatively determine an *answer* to the question of what to do, that I cannot *ask* the question. For there is such a thing as being in a state of uncertainty as to what to do. Indeed, agents in these dilemmas are often our paradigmatic examples of being stuck in practical uncertainty. And it is natural to capture uncertainty as a condition in which one is asking oneself a question that one finds oneself unable to answer.

Intrinsic conflicts have a perspective-dividing quality: seeing one option as valuable gets in the way of seeing the other as valuable. This is, again, because the self-monitoring activity connected to our valuing of one of the options requires us to reject the inclination to be attracted to the other option. In extrinsic conflicts that take the form of a dilemma, we are all too able to see both options as valuable. The problem is that we don't want to choose between them. We really, really want both goods. Thus dilemmas are in a way the opposite of intrinsic conflicts: the latter are characterized by an inability to see both values at once, the former by an inability to do anything *but* see both values at once. The most heart-wrenching dilemma displays nothing more than the contingent unrealizability of two values. It does not differ, in the relevant respect, from the aesthete's conflict between the movie and the concert. What I have called intrinsic conflicts, on the other hand, are conflicts between the values themselves.

Intrinsic conflicts are not resolvable by deliberation because one cannot step back far enough from them to even get the deliberative question in view. They are deliberatively irresolvable by being, in the first place, deliberatively *unavailable*.

V. Intrinsic Conflict and Ethical Theory

In the moral psychological literature, the topic of deliberatively unavailable conflict has not been singled out for special treatment or analysis; but it has been doing important philosophical work behind the scenes. The fact that some conflicts are structured so as not to be amenable to reflection has formed the basis for arguments for or against Aristotelian, Kantian, and consequentialist theories⁶⁵. I offer three examples.

⁶³ Chang adds: on par with, i.e. roughly equal. But both 'equal' and 'on par' are problematic for the agent trying to make a decision. Sophie might feel is that she is in no better position to choose between her children if she decides that their lives are of equal value (or on par), by comparison with the case in which she decides that she cannot determine the value relation in which they stand.

⁶⁴ Though in this case I might be in a position to 'just pick.' See above, p.37.

⁶⁵ Richard Holton (2004) also adverts to the phenomenon as part of his case for the distinctive (and perhaps irreducible) normative status of resolutions. He identifies a intrinsic conflict between, on the one hand, the value of holding fast to what one has resolved to do, and, on the other hand, the value represented by the temptations whose influence the resolution was formed to resist. So, Homer resolves on Friday to go jogging on Saturday morning even if he feels tired then. Come Saturday morning, Homer cannot weigh the value of sticking to his resolution against the value of assuaging his tiredness. For part of what it is to appreciate the value of being resolute, in this case, is to deem tiredness an irrelevant consideration. Holton takes it that when Homer resolved, on Friday, to go jogging on Saturday despite tiredness, he resolved that he would not, on Saturday, consider tiredness as a reason to avoid jogging. Homer cannot, on Saturday, weigh the value of resoluteness against the value of sleep because asking this question entails already having abandoned the resolution. He can, of course, still weigh the considerations on the basis of which he has resolved—e.g., the fact that running promotes health—against the value

John McDowell has criticized a picture of virtue on which “the virtuous person’s judgment is a result of balancing reasons for and against.” (p.55) He offers an Aristotelian alternative: the virtuous person should not be understood as judging that, on balance, the reasons to act well outweigh the reasons to act badly. Rather, the reasons to act badly are simply not practically salient to him: “some aspect of his situation is seen as constituting a reason for acting in some way; this reason is apprehended, not as outweighing or overriding any reasons for acting in other ways....but as silencing them.” Whoever experiences as salient the reasons to desert his comrades has already, according to McDowell, put himself out of the running for courage. Such a person cannot (fully) redeem himself by going on to prioritize other, better, reasons. The virtuous soldier is one who just doesn’t experience reasons to desert as pressing or salient in the first place.

Bernard Williams makes a similar point about the conflict between partial and impartial values. Williams thinks that the natural way for a husband to resolve the question of which of several drowning people to save is simply to see that one of the drowning people is his wife. But Kantianism, according to Williams, requires the person to take into account the claim of each drowning person to be saved. Though the Kantian may allow the husband, in the end, to privilege the claim of his own wife (“in situations of this kind it is permissible to save one’s wife”), Williams finds it objectionable that the husband was required to deliberate in this way: “such things as deep attachments to other persons will express themselves in the world in ways which cannot at the same time embody the impartial view.” (1981, p.18) The point of view on value in which one impartially considers *everyone’s* demand to be saved is incompatible with the one from which one person stands out as a special locus of importance. In forcing the husband to ascend to a level of deliberation that can include the value of others, the Kantian forces him to abandon his love for his wife. Williams, then, thinks that love, at least in some cases, conflicts *intrinsically* with Kantian Morality.

David Sobel⁶⁶ deploys deliberative unavailability as an objection to consequentialism. A consequentialist defines the best choice as the choice of the option with the most value, and she is committed⁶⁷ to there being a fact of the matter about which option this is. One way to spell out that fact is to refer to the putative choice made by an ‘ideally informed agent’ familiar with all the options in question. Sobel argues that the “the notion of a fully informed self is a chimera,” (p.794) because the great variety of possibilities, choices, and lives an agent might lead are not available to a single consciousness. Consider, for instance, the difficulty of combining in one consciousness the point of view of a life characterized by certain forms of ignorance as to other lives (e.g. that of an Amish person), and the very lives of which such a

of satisfying his desire to sleep. But to open up that deliberation is to abandon the prior work that he did in resolving.

⁶⁶ I’ve chosen Sobel as representative of a family of views objecting to consequentialism by objecting to the possibility of full-informedness. See also Rosati 1995, Velleman 1998, Anderson 1997.

⁶⁷ Peter Railton distinguishes himself from other, “overly optimistic” utilitarian in disavowing this commitment (p.735-6). He suggests that a more modest consequentialists might simply concede that there is a certain amount of indeterminacy as to what someone ought to do. Sobel responds that this concession limits the force of the consequentialist injunction to maximize value (fn. 4).

person must be ignorant⁶⁸. Sobel's argument for the impossibility of such an agent rests, then, on the claim that some of the values or points of view she would have to 'reckon up' represent intrinsically conflicting points of view on value.

The intrinsic conflicts to which these authors refer us do not amount to psychologically real events occurring in a person. If the reasons of the coward and the reasons of the brave man belong to, or in, two different people, it is not clear that any single person can be torn between them. Williams' famous phrase, "one thought too many," indicts the husband for having as much as a Kantian *thought*. If the suggestion is that in a given situation one *either* inhabits the value-space of Kantian morality, *or* one sees the world as a husband does, then there is no being 'torn' between these outlooks. Sobel, likewise, describes the cases which interest him as those in which the very access to the value of one life might preclude another life from being "available to one's consciousness." (p.794) None of these philosophers is committed to the claim that a conflict from which one cannot step back can nonetheless be experienced.

On the other hand, none of them explicitly *denies* the possibility of such a conflict. And we have seen that it is possible, at least in some cases, to reconcile the presence of two incompatible ethical perspectives: we can understand them as placing competing demands on a subject's ethical attention. In this way we've defended a Frankfurtian commitment to the possibility of the *experience* of intrinsic conflict, such as being torn between love and strife. For Frankfurt insists that even after 'identifying' with love and 'externalizing' spite, such an agent might still *feel* both love and spite⁶⁹. The fact that they do not consider the possibility of psychologically real intrinsic conflicts means that McDowell, Sobel, and Williams are not moved to consider how one might resolve such a conflict. For us, as for Frankfurt, this is an important question.

VI. Two Suggestions for Resolving Intrinsic Conflicts

(a) By Identification (Frankfurt)

Intrinsic conflict cannot be resolved by deliberation. How, then, can it be resolved? On Frankfurt's view, an intrinsically conflicted (or "ambivalent") agent "decides" to cut off one of the two desires, "identifies" with the other, and thereby "constitutes himself" as wholehearted (1988, p.170). After having made such a decision, "...the person no longer holds himself apart from the desire to which he has committed himself. It is no longer unsettled or uncertain whether the object of that desire—that is, what he wants—is what he really wants..." (1988, p.170) Frankfurt depicts the agent as deciding between conflicting desires by standing back ("holds oneself apart") from them and somehow plumping for the one over the other ("mak[ing] up one's mind," 1988, p.172). It is at this point that we see the importance of articulating the precise *way* in which intrinsic conflicts are deliberatively irresolvable. For it turns out that the

⁶⁸ Sobel's argument ranges over a wide variety of examples; I've fixated on the example of Amish vs. Epicure for the purposes of brevity.

⁶⁹ "...he may continue to experience the rejected desire as occurring in his mental history..." (1976 p.250), "when someone identifies himself with one of his desires, the result is not necessarily to eliminate the conflict between those desires, or even to reduce its severity, but to alter its nature....Quite possibly, the conflict between the two desires will remain as virulent as before." (1988, p.172)

‘decision’ to which Frankfurt adverts here is appropriate only to resolving those conflicts where the deliberative question can be so much as posed.

Consider, again, the many ways in which extrinsic conflicts may not be susceptible to deliberative resolution. Sometimes the agent does not have the time to do the requisite deliberating; or lacks crucial information; or finds her options to be of exactly equal value. In any of these cases, the agent can resolve the conflict by simply picking⁷⁰ one option instead of the other. This same method is available for the agent faced with incommensurable values: she can just pick. Of course, this is no guarantee that she *will* pick. Perhaps she is stymied by uncertainty long enough that the opportunity for choice simply passes, and she is stuck with one (or neither) of the options she couldn’t choose between.

The point is that deciding for the one option over the other is a method of resolving extrinsic, and not intrinsic, conflicts. For the idea of a decision already invokes options being presented, as it were, side by side. Someone can make a decision between two things only if she can, in all seriousness, ask herself which of the two she should do. Recall our three extrinsically conflicted variants of the wife—the one who is teaching her husband a lesson, the one who knows the letter contains damaging revelations, the one who is running for office. These people can ‘step back’ and decide the conflict in terms of, respectively, love, hate, or political ambition. If the wife who *is* intrinsically conflicted tries to “step back” far enough from loving her husband that the value of causing him misery comes into view, she’s stepped out of the attitude of love altogether.

Frankfurt imagines that the intrinsically conflicted agent who has not ‘resolved’ his intrinsic conflict is “uncertain which side he is on, in the conflict between the two desires” (p.172) and that he is looking to resolve this uncertainty. But uncertainty is, in fact, a mark of extrinsic conflict, in particular, those extrinsic conflicts that constitute dilemmas. The moviegoer might be so heavily invested both in seeing this movie and in attending this concert that he finds himself wavering between the two options, unable to decide which to do. He might say to himself, “I do not know whether I really want to attend the movie, or whether I instead I prefer the concert.” He might, then, use deliberation as a way of resolving this uncertainty. Agents who are intrinsically conflicted are, by contrast, *not uncertain* as to which side of the conflict they are on. If there is no question an agent seeks to answer, she cannot be described as uncertain.

Because the intrinsically conflicted agent cannot step back and get both options in view, her conflict takes an asymmetrical form. Frankfurt describes an intrinsically conflicted agent thus: “Suppose that a person wants to compliment an acquaintance for a recent achievement, but that he also notices within himself a jealously spiteful desire to injure the man.” (1976, p.249 emphasis mine) The desire to compliment constitutes, at this moment, this agent’s ‘main’ perspective, one from which the desire to injure represents, at this moment, the irritating (or disturbing) distraction. Such a person does not seriously entertain the possibility that injuring is the right thing to do; he is looking for ways to silence or ignore this impulse. When the desire

⁷⁰ I call this “simply” picking to contrast with decisions that are products of deliberation. I do not mean to suggest that such acts of the will (and their attendant methods, e.g. coin-flipping) are simple to understand. But it is not my task, here, to account for them. See Ullmann-Margalit and Morgenbesser for a discussion.

to injure appears on the scene, a negative judgment has already been entered against it. The way in which such an agent combines love and spite is akin to the way in which we ‘combine’ incompatible demands on our perceptual or cognitive attention: we attend (however briefly) to one of the two, and experience the other as a diminishment or defect in that very attention. We actively try to silence the one demand, and accentuate the other. We may flit back and forth over the course of some stretch of time, but at any given moment, we favor either the one side, or the other.

In theory, it seems that we should allow for the case in which one feels *exactly as much* love as spite in a given moment, but it is in fact hard to imagine what that is like. When an ethical outlook dominates, then there is something the agent is doing, e.g. pursuing her husband’s happiness, and something that is distracting her from doing that, e.g. pursuing his misery. Someone perched on the knife edge of love and hate cannot be described as doing both. Trying to love and hate at the same time is like trying to philosophize and take care of the baby at the same time—under the presupposition that one cannot do both. Such an agent would be frozen, like a deer in the headlights. She would be neither loving, nor hating (nor deliberating between loving and hating, since that can’t be done). If such a state is a psychological possibility, it does not represent a state of uncertainty or reflection. Rather, it is some form of mental disturbance in which the presence of conflicting desires is so extreme as to be valuationally disabling. Without empirical investigation, it is hard to know what to say about this condition, but I suspect we may end up saying that the attempt to feel too many things at once leads to not feeling anything at all. We should not be fooled by the fact that a ‘frozen’ agent—if she exists—may look, from the outside, like one who is stepping back and reflecting. It is important not to confuse a state of distress with a state of inquiry, though both dispose someone to stand still and not do anything. If ‘frozenness’ is possible, it is not typical; and it does not vindicate the possibility of resolving intrinsic conflict by deliberating (or deciding).

Frozenness aside, intrinsic conflict is characteristically asymmetrical. Its resolution must, therefore, take a distinctive form. What one wants, in seeking to resolve a intrinsic conflict, is to fully, undistractedly inhabit a point of view, to act undistractedly, to stop hearing a ‘wrong’ voice. An intrinsically conflicted agent will not necessarily vacillate between her two outlooks—it might be the case that one of them always dominates—but those that do will also look to resolution as an end to this vacillation. Frankfurt misunderstands the problem of intrinsic conflict as one of uncertainty. His solution—identification—doesn’t solve the real problem. Frankfurt acknowledges that “when someone identifies himself with one rather than with another of his own desires, the result is not necessarily to eliminate the conflict between those desires, or even to reduce its severity...” (1988, p.172). Insofar as identification is compatible with the desire remaining exactly as strong—and thus exactly as distracting—as it always was, it doesn’t resolve the agent’s feeling of conflictedness. Nor does identification insulate someone against future vacillation. Suppose the wife ‘identifies with’ her love, but in the next moment, experiences such a powerful upsurge of spite that she is now gleefully driven by spiteful hatred. She cannot, on Frankfurt’s view, still be said to be identified with her love. For Frankfurt denies that there is any *historical* dimension to identification. But if this is right,

then Frankfurt can draw no distinction between the desire I identify with and the desire that holds the psychologically dominant position⁷¹.

Before turning to my solution, I would like to address a worry as to the coherence of my description of intrinsic conflict⁷². I have argued the agent cannot get the conflict into view by occupying a third, more reflective perspective encompassing both of her intrinsically conflicting values. This is because the values tend to annihilate one another: the value-nexus of each contains attitudes that undermine the legitimacy of the other. In that case, one might wonder how it is possible for the agent to so much as register the conflict. Intrinsically conflicted agents are aware of their conflict, and experience their condition as a kind of turmoil they would like to resolve. Doesn't this imply that there is at least some sense in which they occupy a third perspective?

On my account, the conflict between the two values is indeed experienced by the agent, because she registers the non-dominant perspective in the fact that she cannot completely inhabit the dominant one. So, in a loving moment, her spite manifests in an inability to be *completely* loving, in the fact that she must try and struggle to love. Why doesn't her very recognition of this fact constitute a third perspective? It could. This agent could observe herself and note that she is a curious specimen, moved by both love and hate. My point is only that whatever attitude she takes, if it encompasses both of the conflicting values, it cannot be a point of view from which she *deliberates*. She cannot take both the mailing and trashing seriously as options for dealing with the letter in her hand. These options can both occur to her, and they can both motivate her, but she cannot deliberate between them. It is important, in this context, to observe that not everything that motivates us is something that can show up for us in deliberation. So, for instance, if I am self-deceived about my desire to hurt someone's feelings (I think I am just offering her constructive criticism), then I will not ask myself, "shall I hurt her feelings now?" I may well be motivated by a desire to hurt her feelings, but I cannot deliberate as to whether I ought to be so motivated. For a different reason, the person who is intrinsically conflicted can be motivated to mail, or to trash, but she cannot ask herself which of the two forms of motivation is better. She could ask herself whether trashing the letter or burning it and leaving the ashes on his desk is a better way to exact her revenge; or, whether dropping it in the mailbox or delivering it herself is a better way to help him out. By contrast, the options of mailing and trashing cannot cohere into a single deliberative question.

(b) by Aspiration

⁷¹ To argue that identification/externalization doesn't resolve intrinsic conflict is not to expose these concepts as empty. For Frankfurt, as has often been observed, uses them to do a variety of jobs: there is a psychological job, where 'external' means something akin to phenomenologically alien (in the mode of obsessional thoughts). And then there is a normative job, where 'external' means something like 'not arrived at by/sanctioned by my reasoning'. Arpaly and Schroeder point out that I needn't, for instance, experience psychological alienation from any desire that moves me akratically. Nor do I need to be identified with one that moves me enkratically. Each of these two senses of 'externality' comes apart from the other—Schroeder and Arpaly argue that we should prefer an interpretation of Frankfurt on which we privilege the psychological sense, whereas Moran argues that it is better to emphasize the normative sense. The debate between them is untouched by the criticism I offer here, since neither sense is dependent on conceiving of identification as the solution to intrinsic conflict.

⁷² I am grateful to an anonymous referee for raising this objection.

Extrinsic conflicts eventually get resolved. They are usually resolved by the person herself: she decides, either in accordance with a deliberative judgment as to which option is better, or in spite of having formed such a judgment (akratically), or in the face of the impossibility of forming such a judgment (no time to deliberate, incommensurable or equal values, not enough information, etc.). If she delays long enough that the opportunity for choice passes, the conflict is resolved without her resolving it. One doesn't go through one's whole life wondering which of two things one should do. Intrinsic conflicts, by contrast, can be lifelong. Some relationships just are 'complicated,' whether they be with individuals (relatives, friends, bosses), aspects of ourselves (weight, heritage, sexual orientation), larger entities or groups of people (one's country, the institution at which one is employed, the intelligentsia, the media), or even ideas (feminism, relativism, liberalism).

Indeed, some 'complexity', some of the time, seems to be the norm for creatures like us. Resolution⁷³ is rarely completed to a point of McDowellian purity, and twinges or reminders of the rejected point of view may surge up occasionally. So let us ask instead whether a person who is very conflicted in some area of her life can come to be less conflicted in that area of her life. This is, perhaps, all that the wife asks of herself: to be more loving, to be less spiteful. Now it is surely possible that she passively gravitates toward love or spite due to no work of her own. Outside forces can change us, just as they can resolve our extrinsic conflicts. What we want to know is whether there is anything the agent can do to resolve her intrinsic conflict *herself*. Can she come to be less conflicted than she is as a result of her own agency? Is there some act of her will—not deliberation, not identification, but some other option—the performing of which would resolve a intrinsic conflict? Yes.

An agent resolves her extrinsic conflicts by deciding, possibly as a result of deliberating. An agent resolves her intrinsic conflicts by aspiring. Aspiration is the diachronic process by which an agent effects change on her own ethical point of view. Aspirants aim to direct their own ethical attention in such a way as to more fully appreciate one value or set of values, and to become immune or insensitive to those values which intrinsically conflict with the first set. An aspirant is someone who works to improve her desires, her feelings, her ethical evaluations, and, more generally, her own capacity for responding to reasons.

Just as we can distinguish a kick that is willed from a kick produced by reflex, so too we can distinguish two ways in which someone can become less spiteful. A non-aspiring variant of the wife—perhaps she was raised to think that marriages are by nature 'complicated'—might surprise herself by gravitating in the direction of unspiteful love. An aspiring variant of the wife might, by contrast, fail in her project of becoming less spiteful. The aspirant has marked out one attitude as a kind of target toward which she orients herself, and another (or others) as a danger from which she must turn herself away. An agent who 'identifies,' in Frankfurt's sense,

⁷³ There is also a distinction, not important for our purposes, between two forms such resolution can take. The first form is the elimination of exactly one of the two conflicting states. The wife can come to stop hating her husband and love him unreservedly, or vice versa. The second is the replacement of both conflicting points of view with a third. Exclusive attitudes, values and forms of reasoning needn't be exhaustive. So the wife can come to a kind of sad resignation that her husband will never do housework, or make romantic gestures, and cease resenting him for it. Maybe she loves him a little less, but she's also stopped hating him. Or she can decide to rid herself of both love and spite by excising him from her life; her aim is to be in a condition where she never thinks about him again. Indifference may replace both love and spite.

with some desire constitutes herself as being in some condition at the time at which she so identifies. She fixes her own identity, statically, by deciding. The aspirant stands in a dynamic relation to ‘her true self’; aspiring does not make it the case that she *is* her true self. Aspiring is the method by which she works to become her true self.

Since the problem of conflict is a problem of divided ethical attention, aspirants attempt to exercise their agency with respect to their own attention. This is true of intrinsically conflicted agents more generally, as we noted above. They ignore those facts, people, or mental images, attention to which would bring to light one (distracting) set of reasons, and they focus on those associated with another, incompatible (dominant) set of reasons. Intrinsically conflicted agents interfere in their own thinking; intrinsically conflicted *aspirants* interfere for the sake of effecting long-term change. In the next chapter we will have occasion to look more closely at this difference, by examining a non-aspirational species of intrinsic conflict. For now, we can say: a conflicted wife tries not to look at the trash can to ensure that she mails the letter; a conflicted and aspiring wife tries not to look at the trash can both in order to mail the letter and in order to become less spiteful. Her goal is to handle these conflict-situations in such a way as to eventually mitigate their occurrence. She directs her attention with the aim of giving it a new shape.

As in the case of the two kicks, skeptical worries will arise. Someone might doubt whether there really is anything *to* aspiration over and above someone’s ending up with a different character. A skeptic might deny that there is a principled difference between ending up less spiteful as a result of *willing* that state, and ending up less spiteful through accident or external influence. In chapter 5, I argue that it is, in fact, possible to exercise one’s will with the aim of becoming a different—better—person. My interest here has not been in defending the possibility of willfully becoming someone any more than in defending the possibility of willfully kicking. My interest has been to articulate the difference between becoming and kicking. So it is perhaps safer, until we have the argument of chapter 5, to state my conclusion in the form of a conditional: *if* there is any solution to intrinsic conflict, it will have to take the form of aspiration.

But how fundamental is the difference between aspiration and identification? A supporter of Frankfurt might suggest that we reconstitute identification as deciding to aspire, or as making up one’s mind as to what to aspire to. This will not work. First, because *deciding* to aspire doesn’t resolve anything. What resolves conflict is aspiration itself, the temporally extended work of changing ourselves, our values, our desires, our outlook. The problem with Frankfurtian externalization is that there is nowhere for the rejected-but-still-existent desire to go but back into oneself. A desire must be *someone’s* desire. Aspirants aren’t done until they have “externalized” the desire all the way into non-existence.

Second, aspiration is not always or even typically heralded by a *decision* to begin engaging in it. Not that it can’t be: perhaps the wife ‘decides to aspire’ at the revelatory moment when she first sees her husband holding their newborn child. Until that point, she has acquiesced in their ‘conflicted’ relationship, but the sight before her occasions in her a resolution to become a different kind of spouse. But that needn’t be how it goes. Quite often we simply find that we *do* aspire to become more loving and less spiteful in a given relationship, and that we have so

aspired for as far back as we can remember. It doesn't undermine the status of the aspiration as the work of the agent's own will that there was no moment at which she "made a decision" to engage in it. Likewise for the case of ceasing to aspire: it can be a matter of consciously, suddenly, giving up, or one can give up little by little, without realizing it, over a long time. The presence or absence of a decision seems to make no difference to cases like these. And that is because they concern a form of willing that does not happen in a moment, but over long stretches of time.

It is a theme of Frankfurt's work, early and late, to reject genealogical explanations in favor of what he calls "structural" explanations. He wants, for instance, to resist the idea—which he associates with Aristotle—that an agent could be responsible for what she does at one time in virtue of something she did at another, earlier time⁷⁴. What matters, according to Frankfurt, is what the agent wills or identifies with *at the time of action*, and not earlier. And so he endeavors to capture the ethically salient features of personhood in terms of relations in which the agent stands to herself (her desires, motives, commitments) at the time of action. But some of the relations in which an agent stands to herself might be essentially historical. To be an aspirant is to proceed away from some (rejected) past set of desires, commitments, etc. toward a different, anticipated future desire set of desires, commitments etc. Aspiration is a form of ethical movement, and nothing can move in a moment.

Aristotle had no notion of the will⁷⁵. Those of us who do have some tendency to understand it as something that acts ("decides") instantaneously. But the will, like the agent herself, has a variety of tasks, and some of them take longer than others. Frankfurt was on to something important when he pointed us to a different *kind* of conflict from the familiar concert/movie kind. And he was right that the resolution of this other kind of conflict presents a new and exciting philosophical problem. But he did not fully grasp how new and how exciting this problem is. Intrinsic conflicts reveal the diachronic work of the will, and they do so because they cannot be resolved in a moment.

When we are extrinsically conflicted, we are able to survey the values between which we are torn from a reflective distance. What happens next can take a variety of forms: we may deliberately resolve the conflict, we may decide without deliberating, we may find ourselves stymied by indecision. When we are intrinsically conflicted, no such synoptic position is

⁷⁴ "This suggests another respect in which Aristotle's theory is unsatisfactory. He maintains that a person may be responsible for his own character on account of having taken (or having failed to take) measures that affect what his habitual dispositions are. In other words, a person acquires responsibility for his own character, according to Aristotle, by acting in ways that are causally instrumental in bringing it about that he has the particular set of dispositions of which his character is constituted. I think that Aristotle's treatment of this subject is significantly out of focus because of his preoccupation with causal origins and causal responsibility. The fundamental responsibility of an agent with respect to his own character is not a matter of whether it is as the effect of his own actions that the agent has certain dispositions to feel and to behave in various ways. That bears only on the question of whether the person is responsible for having these characteristics. The question of whether the person is responsible for his own character has to do with whether he has taken responsibility for his characteristics. It concerns whether the dispositions at issue, regardless of whether their existence is due to the person's own initiative and causal agency or not, are characteristics with which he identifies and which he thus by his own will incorporates into himself as constitutive of what he is." (1988, pp.171-2)

⁷⁵ See ch. 2 of Frede for discussion.

available to us. Instead of stepping back to a neutral point of view, aspirants step forward—little by little—into a state of lessened conflict.

When we aspire we resolve intrinsic conflicts gradually, over an extended period of time. We cannot, however, resolve all intrinsic conflicts in this way. In the next chapter, I will describe a species of intrinsic conflict that is not amenable to aspirational resolution precisely because they are by nature short-lived. I hope thereby to shed light both on the diachronic quality of aspirational conflict resolution, and on the species of intrinsic conflict in question—for it is of substantial independent interest.

* Chapter 4: Akrasia

It is a common human predicament to find oneself eating or imbibing or smoking or sleeping more than one thinks one should. We judge that we have most reason to refrain from indulging, and yet we willingly indulge. This phenomenon is known as ‘akrasia’ or ‘weakness of will’ or ‘incontinence.’

In the twentieth century, Donald Davidson ignited philosophical interest in akrasia by offering a new characterization of the phenomenon in his paper “How is Weakness of the Will Possible”. Davidson criticized his predecessors⁷⁶ for moralizing akratic conflict by forcing it into the mold of reason vs. passion. The akratic who is moved to sensual indulgence against the dictates of temperance or courage or justice is, he notes, only one kind of akratic. Davidson points out that akratics can act on (and against) just about any kind of reason, offering as an example someone who akratically rises from bed to brush his teeth against his considered judgment that it would be best to stay in bed and go to sleep. Davidson presses us to recognize anyone who acts against her better judgment as akratic.

The akratic is someone who fails to do what she takes herself to have most reason to do. This is not to say she acts a-rationally, for no reason at all. She acts on *a* reason, but she fails to act on what she takes to be her best reason. (See Davidson 1980, fn. 25) Davidson (and many others) have sought to offer an account of akrasia that leaves it open what *kinds* of reasons these are. His insight lies in his observation that the authority of the judgment the akratic acts against is not *moral* but *deliberative*. The akratic’s action is wrong not because she (necessarily) violates any moral principle, or yields to some base passion. Her sin is irrationality as such—the violation of the deliberative procedure embodied in what Davidson calls her “all things considered judgment”. In this chapter, I will bring out how the theorist of aspiration is in a uniquely good position to fill out Davidson’s demoralized conception of akrasia.

I will argue that akrasia is the result of a special kind of intrinsic conflict arising from the mechanics of the deliberative procedure itself. Sometimes, I reason about how best to act *even though I already have a pre-deliberative answer to that question*. For instance, it seems to me that I should get out of bed and brush my teeth. Instead of acting on that impulse, I check it and reflect. I set aside my immediate inclination to rise long enough to ask myself “what ought I, really, to do?” If, on balance, the reasons seem to favor staying in bed, then I am torn between my two answers as to what I ought to do: the answer that I arrived at immediately and without reflecting, and the answer that I arrived

⁷⁶ Plato, Aristotle, Aquinas, Butler, and Mill come in for special criticism, but he also generalizes the point: “I know no clear case of a philosopher who recognizes that incontinence is not essentially a problem in moral philosophy, but a problem in philosophy of action.” (fn. 14)

at via deliberation. These two answers embody the conflicting evaluative perspectives from which an agent can go on to be either akratic or enkratic (strong-willed).

The standard approach to akrasia takes for granted the absence of intrinsic conflict by presenting the reason on which the akratic acts as deliberatively outweighed by her better reason. As I argued in the previous chapter, an agent can see one reason as outweighing another only if the two reasons are contained within a single evaluative perspective. In the opening of this chapter, I explain the problem with confining the akratic's conflict within a single evaluative perspective: one is forced to describe the akratic as acting on her acknowledgedly weaker reason. In another paper⁷⁷, I argue that it is impossible to act on a reason that has been outweighed in one's deliberations. Here, I restrict myself to pointing out that even if such a phenomenon were possible, it would amount to a much more serious form of rational defect than akrasia.

The large literature spawned by Davidson's paper converges on little other than an acceptance of the assumption that akratics act on a reason they acknowledge as weaker than another they could have acted on⁷⁸. If I am right that this assumption is the problem, and that the key to avoiding it is to locate the akratic's two reasons—the one she acts on, and the one she (thinks she) should act on—in different evaluative perspectives, why haven't the many theorists of akrasia lit upon the presence of intrinsic conflict?

To some degree they have: Frankfurt's discussion of different 'orderings,' alienation and externalization is motivated, in part, by a desire to explain akratic conflicts⁷⁹. But the investigation of akrasia is not the easiest road into a theory of intrinsic conflict, because akratic intrinsic conflicts, unlike most intrinsic conflicts, have a short duration. The akratic's intrinsic conflict begins when the agent arrives at a considered judgment conflicting with her original inclination, and ends as soon as she acts⁸⁰. We will have an

⁷⁷ Callard, "The Weaker Reason".

⁷⁸ Walker brings this out in his survey of accounts of akrasia; I update his summary with some additional references in fn. 3 of my paper, "The Weaker Reason."

⁷⁹ See Schroeder and Arpaly, who argue that Frankfurtian externalization/alienation conflates a feature of one's will and a feature of one's phenomenology. Only the former would pertain to akrasia (See above, fn. 71). See also Davidson's 'Deception and Division,' in which he returned to the topic of weakness of will later in life. There he argues, drawing on Freud's conception of a divided mind, that the akratic's weaker reason, and her all things considered judgment occupy different mental partitions. The way in which the view I describe here differs from Davidson's is that Davidson marks the partitions solely on the basis of the fact that they happen to contain different reasons. I take the akratic to be divided not only between reasons but between forms of reasoning. Like Freud, I think that mental partitions must mark differences in kinds of mental activity, so that the lines of division are divisions between ways of thinking. A geographical picture of the mind is suggestive, but representing the relevant 'regions' as different evaluative perspectives helps to cash out the metaphor. Davidson does not offer any account of how the relevant sets of reasons are separated from one another; my account of intrinsic conflict in chapter 3 above was an attempt to provide just such an account.

⁸⁰ Akratics may experience such conflicts in a recurrent or regular way, but that possibility isn't written in to the definition of the phenomenon as 'acting against one's better judgment.' There is no conceptual problem with the idea of a singularly tempting object or circumstance (recall Sue, p. 46 above). Indeed, even in cases of recurrent temptations (a fondness for sweets that regularly thwarts one's attempts to diet) the disposition is merely recurrent, rather than deeply structuring a wide variety of responses. Action

easier time getting a handle on the phenomenon of conflicting evaluative perspectives if we take our bearings from cases such as the bitter wife. For in that sort of case we can often use the history of the agent's psychological development to identify the two longstanding threads of emotional responsiveness that we find tangled in a given moment of, e.g. turbulent letter mailing or letter trashing. It will be easiest to pick out these threads if the case has an aspirational trajectory: if we can identify hate and love as endpoints of an agent's process of self-change, it will be relatively clear that the conflicted emotional state that we find at some moment along the way represents an intrinsic conflict between the point of view the agent has, and the one she wants to have. Not all intrinsic conflicts are aspirational, but someone who is attending to aspiration is especially well placed to get intrinsic conflict in view.

In Plato's *Republic*, Socrates proposes that they investigate justice in the city, since that will make it easier to see justice 'writ small' in the soul. The standard approach to akrasia is handicapped by the simple fact that it approaches the phenomenon on its own terms, rather than as a miniaturized variant of the kind of conflict that we see in its fullest or clearest form elsewhere. The theorist of aspiration can assist the theorist of akrasia by supplying the larger context in which the phenomenon becomes more readily intelligible. And, as in Plato's *Republic*, illumination runs both ways: as a case of momentary intrinsic conflict, akrasia brings into relief the importance of temporal extension to aspiration.

I. Acting on a weaker reason?

In virtue of what does the judgment that the akratic acts against qualify as her *better* judgment? The judgment cannot be better because of the kind of value it promotes, if we want to allow for the possibility of being akratically prudent or perhaps even akratically moral. Instead, Davidson and others⁸¹ claim that the judgment owes its authority to the fact that it has taken account of the reason the agent acts on. This reason is sometimes called her "weaker reason." The reason on which an akratic acts is termed "weaker" because it is outweighed in the deliberation of which her better judgment is the upshot. The, Davidson calls the agent's better judgment her "all things considered" judgment. Having compared her reason to indulge with her reason to refrain, the akratic judges that "all things considered" she should refrain. The strength of such a judgment lies in what we might call its *inclusiveness*, which is to say, in the fact that one of the things considered in it is the reason to act as the akratic goes on to act. While her weaker reason says, "I should eat, because the cookie would be tasty," her all things considered judgment says "I should refrain from eating, even though it would be tasty."

In "The Weaker Reason" (2015), I argue that it is *impossible* to act on one's weaker reason. If one has discounted the rational force of one consideration in favor of another, one cannot then take the first consideration as a reason to act. If the reason the akratic acts on has been counted in her deliberations, it cannot be counted again, independently,

theorists working in the wake of Davidson take "akrasia" to refer to a kind of action rather than a kind of character trait.

⁸¹ See my, "The Weaker Reason" for references and further discussion.

as a reason to act. I argue that on the standard account of akrasia, the akratic emerges as engaging in an unintelligible form of double counting, instead of an eminently intelligible form of self-indulgence. Without presenting the full argument here, let me try to convey some of its force by introducing the an absurd case that fits the description, ‘acting on one’s weaker reason.’

Offered a choice between \$100 and \$200, J chooses \$100. She needs money, and there is no other difference between the two offers. When asked why she did that, she says “\$100 is still good—I had *some* reason to pick \$100, I just had *stronger* reason to pick the \$200. I acted on my weaker reason.”

In these kinds of circumstances, we are not inclined to think we have a case of akrasia on our hands. If we can accept the case as described, we will do so by diagnosing J with some more substantial form of rational breakdown than akrasia. The problem seems to be the fact that she has exactly the same sorts of reasons to choose the \$100 as she has to choose the \$200. Since they are identical apart from amount, any reason J has to take the \$100 has already been accounted for in her judgment that it would be better, overall, to take the \$200. The \$100 still has value, of course, but this value cannot, in this context, serve as J’s reason to choose it. For the value has already been tallied in the reasoning that concludes in the judgment that it would be best to choose the \$200. Insofar as she’s moved to acquire money, it makes *no sense* at all that J would choose the smaller amount.

Akratics don’t, of course, look like J. They find themselves faced with heterogeneous rather than homogeneous choices. It is intelligible that someone manages to eat the cookie even after judging that she will be happier, overall, if she sticks to her diet. In that kind of case, we can see how the appeal of the akratic action might survive the judgment that it would be better to do something else. But Davidson and the inheritors of his ‘weaker reason’ conception of akrasia cannot exploit this difference between the two cases so long as they retain their commitment to locating the authority of the akratic’s ‘better judgment’ in its inclusiveness. For the heterogeneity in the two values can only be relevant to the akratic choice insofar as it presents deliberative obstacles that impede the formation of an inclusive judgment.

Suppose that someone, having been offered an especially delicious cookie, knows that she will feel immense self-satisfaction later if she sticks to her diet; but she also knows that she will really enjoy the taste of the cookie. Until she has reckoned which of these two values outweighs the other, she hasn’t formed the kind of judgment against which she might act akratically. If, for instance, she simply cannot say whether self-satisfaction outweighs pleasure, or vice versa, she leaves the issue of ‘best overall action’ undecided. If she eats the cookie, she has not done so akratically—for she did not form a ‘better judgment’ in the relevant sense.

The phenomenon of value heterogeneity certainly might give rise to a situation in which the agent failed to take into account some feature of the appealing action. Perhaps she only noticed how tasty it was at the last minute, and was pressed to make an immediate decision (the cookies were about to be taken away!). Perhaps she couldn’t make a ‘total

accounting' because the value of the cookie—taste—and the value of the diet—health—are incommensurable. In any of these cases her "better judgment" will not be better in the relevant sense: it won't be *inclusive* of her weaker reason. To the extent that the better judgment leaves out the value of the weaker reason, Davidson cannot count it as having been akratically violated. To the extent that the distinctive value contained in the 'weaker reason' is fully accounted for, heterogeneity in value does not help explain how someone might act on a reason.

What the standard view cannot explain is *why* the fact that our akratic options don't look like J's options is relevant to our understanding of akrasia.

II. Akratic Conflicts Are Intrinsic

We can get some purchase on the distinctive nature of akratic conflict by availing ourselves of the apparatus of intrinsic and extrinsic conflict described in the previous chapter. When I judge that I would be happier overall if I refrained from eating the cookie, how is it that I 'take into account' the value of eating it? I might 'account for' its deliciousness by taking the chocolatey smells and the craggy bumps and the warmth radiating from its surface as indications that the cookie will provide an excellent taste experience. But this reflective and evidential treatment of the cookie's features is a different kind of response to those features than the one I have when I smell the smells and see the craggy bumps and feel the warmth and find myself recalling the memory of similar cookies eaten when still warm and melty⁸².

The attitude one adopts to the cookie when one aims to compare its value to that of resolute dieting is different in kind from the attitude one adopts to the cookie when one is savoring the smell of the cookie. If I set aside thoughts of dieting, I can inhale the perfume of the cookie and devote myself to relishing the upcoming chocolate delight. The akratic finds herself thinking in two very different ways about the same phenomenon. The two values are not incommensurable or incomparable, as they might be for someone who cannot decide which she wants more. The akratic's problem is that the relevant values don't show up for her within a single value-perspective. She cannot ask, let alone answer, the question as to which is preferable.

There is a difference between being moved by a feeling and taking the fact that I have a certain feeling as a reason to do something. When I assign *the fact that* I 'feel like' doing something a value, I adopt a meta-attitude toward my feelings that differs significantly from the attitude of simply *feeling* that feeling. Consider, for instance, that I can be moved by *the fact that* someone else feels something, but I cannot be moved—at least not

⁸² Sarah Buss (ms.), in the context of raising an objection to constitutivist accounts of agency, describes the former response as alienated. Buss' contention is that if there is a substantive, constitutive aim of agency, then an agent cannot inhabit the point of view of her own pre-deliberative seemings. Such an agent is alienated from her own motives, estranged from the way things rationally seem to her, because she cannot be responsive to reasons simply insofar as she experiences a state of affairs as desirable. (For she can only take these reasons as reasons by judging that they serve the constitutive aim.) The self-alienation which Buss takes to doom constitutivist theories of rational agency is remarkably akin to the one I locate in the akratic's relationship to her pre-deliberative seemings.

directly—by someone else’s feelings. At best, if I empathize with him, feelings in him can occasion feelings in me.

One may, of course, tell oneself that one has ‘taken into account’ the pleasure of eating the cookie. One may say that one has fully incorporated that value into the judgment that one ought, rather, to refrain. But consider a facially similar claim that might be made by someone choosing \$200 instead of \$100. He, too, might say he has ‘taken into account’ the value of the \$100 in his decision to opt for the \$200—and here, we might add, he really means it. His decision to choose the \$200 leaves nothing out. We cannot imagine what reason he would possibly have for choosing \$100. I want to suggest that the reason why we cannot construct a case of akrasia around the choice between \$100 and \$200 is that we cannot imagine someone who is intrinsically conflicted between these two options. But what, exactly, is the ‘left out’ in the cookie case?

The akratic’s judgment that it is best to refrain took into account the fact that the cookie is tasty. But what he has left out of his prudential calculations is the fact that the cookie *seems* tasty; that it *seems* like he should eat it. The cookie has a delicious chocolaty smell, its craggy bumps look so appealing, it is *right here* in front of me—these all strike the akratic as reasons to eat it. The enkratic is able to ignore these reasons in favor of the reason promoted by her all things considered judgment; the akratic acts on them. But the akratic and the enkratic have an important common ground in the experience of intrinsic conflict that precedes their bad (akratic) or good (enkratic) action. It will be helpful to have a way of referring to the person who is conflicted in such a way as to be either akratic or enkratic. If I call the person undergoing temptation “pre-akratic” or “pre-enkratic” I am deciding the issue of which way she will act, so I will simply call her “pre-kratic.” Many pre-kratics will attempt to distance themselves from the object of their desire: to stay out of the room with the dessert table, or, if they can’t, to avoid looking at the cookie plate, or, at the very least, to avoid inhaling the chocolaty scent. They avoid encountering sensible features of the object in order to deprive their akratic evaluative perspective of the rational resources it needs to engage action. They want to stop themselves from seeing things akratically—to blind themselves to the reasons the akratic ends up acting on.

Sensible features of the environment may, of course, play a deliberative role and thereby be included in the agent’s better judgment. If I am genuinely uncertain whether I will enjoy eating this cookie, my senses might supply me with *evidence* for judging that the cookie is delicious, but even in this case they do not serve as independent reasons for eating the cookie. The fact that the cookie *seems* delicious will not strike me as a reason to eat it, insofar as I have already acknowledged the rational significance of the fact that the cookie *is* delicious. Having concluded that it is delicious, I can no longer make any rational use of the cookie’s smell, its proximity, and its appearance in my project of deciding whether it’s better, all things considered, to eat the cookie or stick to my diet. Indeed, those sensible features are not only irrelevant but distracting.

Consider a case in which the smell is so overpowering that unless I eat one of the cookies, I will spend the rest of the party being distracted by it, and unable to sustain a

conversation. I may judge, on the basis of this fact, that, all things considered, I should eat the cookie. Notice, however, that I am not treating the smell *itself* as a reason to eat the cookie. I'm treating the fact that I have this smell-experience, taken together with the fact that smell-experiences of that kind conduce to anti-social behavior, as a reason. The reason the akratic acts on does not need to be routed through the acknowledgement of implications of the psychological fact that that he's having a certain experience—he's moved by the rational force of the experience itself.

My claim is that in the cookie example, the appearance of the akratically desired object supplies us with reasons to eat it, and that these reasons are of the right sort to motivate an akratic, since they do not—for reasons I will explore below—get 'weighed' or factored into the judgment that it would be better to stick to one's diet. One might wonder whether this account will be broad enough to cover all cases of akrasia. We are not always akratically attracted to (or akratically repulsed by) features of the world to which we have perceptual access. For instance, in an alternate version of the cookie example, I might learn there is a cookie in a neighboring room, and akratically enter the room and eat the cookie. If I acted akratically not only in eating the cookie but already by entering the room—before entering, I judged that it would be best to avoid the cookie room—then my reasons for acting akratically cannot rely on a sense-experience. By the time I can smell or see it I have already acted akratically.

We must allow that one—indeed, a central—way in which the appearance of something can present itself to us is via the imagination. Imagining a delicious cookie can have the effect of seeing one, since imagination is a way in which we make present to ourselves an otherwise unavailable sense-experience. If I cannot see something, I can imagine what it would be like to see it, and thus “see” it in another way. Consider the judgment “the party will be fun,” as incorporated into the deliberations of a pre-kratic who is trying to decide whether to go to the party or to stay home and study. Such a judgment is a very different kind of reasons-response from the one she has to the image in her mind's eye...of a door opening into the light of a room filled with people dancing, laughing, sharing food. One can respond to the convivial warmth of this imagined scene in much the way that one responds to the wafting odor of the chocolate. In deliberating over whether the value of attending the party (the fact that it is fun) outweighs the value of staying home and studying for her test, she does not take into account the reasons supplied by her imaginative access to the fact that the party in her mind's eye *appears* to be fun.

Even in cases of akrasia in which we do have perceptual contact with the object, the imagination is an important source of akratic reasons. For the smell of the cookie may provide some reason to eat it, but the imagined taste—imagined all the more vividly, perhaps, due to the smell—contributes heavily to the agent's sense that the cookie is, as philosophers say, to-be-eaten. For what the agent needs a reason to do is *eat* the cookie, and so it is *eating* the cookie that must strike her as appealing. For this, she needs her imagination, since eating the cookie cannot present itself to her senses until the time for reasons is past. By imagining ourselves engaging in the action in question, we gain access to a set of appearance-based reasons to act. These reasons stem from a perspective

on the action that is unavailable from the point of view of the project of deliberating as to what it's best to do overall.

Indeed, they are so unavailable that there will be some temptation for the agent to think that they are not even reasons. In an intrinsic conflict, when we view the one set of reasons from the point of view of the other, we are inclined to say 'those are not reasons at all.' The bitter wife, in a more loving moment, denies that her husband's behavior gives her any reason at all to throw his letter in the trash. From the ethical perspective of love, reasons of hate don't appear to be reasons.

We saw in the last chapter that intrinsically conflicted agents cannot adopt a deliberative perspective that admits of both values. If such an agent does deliberate, she will be blind to the appeal of one of the two options before her. As an illustration of such blindness, consider a literary example. Zafar, the protagonist of Zia Haider Rahman's novel "In the light of what we know," was born in Bangladesh but raised in London from early childhood. As a pre-teen he is sent back to Bangladesh to learn something of his origins. He describes his arrival in Bangladesh:

Spread along the platform was a mass of bobbing black hair like a long wave of silk. Suddenly I felt the first stirrings of what I would later come to recognize as kinship, a feeling that alarmed me, a sense that I was of a piece with a group of people for the most basic reasons, simple to the senses and irrational. They all looked like me.

Zafar is "alarmed" by a feeling he cannot name, an "irrational" sense that he belongs with or is a part of the strangers around him. His rational judgment says that he is an alien interloper to this place. Though he was born there, he cannot remember it and does not consider it 'home.' Had Zafar articulated this judgment more fully, he might, given his character's intellectual bent and multicultural upbringing, have claimed that where you belong depends not on how you look but how you think. But he responds in a visceral way to the sight of people who look like himself, *as though* belonging were a matter of racial similarity. He evidently takes the fact that people look like him as a reason to consider them 'his own.' For notice that, although he disavows this impulse as "irrational" he also describes himself as moved by "the most basic *reasons*." These reasons are, in Rahman's beautiful phrase, "simple to the senses."

What strikes him, from one ethical outlook, as a reason to consider himself 'home,' strikes him, from another ethical outlook, as irrelevant to that question. Zafar is intrinsically conflicted, and the conflict proves to be a long-term one: questions about the rationality of his feelings of homelessness resound throughout his adult life. Akritic 'reasons' to eat the cookie, stay in bed past the alarm, skip studying to attend the party, accept 'one more for the road', smoke a last cigarette are bound to sound childish to an adult ear: it looked tasty, it felt cozy, it seemed fun, etc. Speaking metaphorically, we might say that the inner adult, having weighed the tastiness of the cookie against the value of sticking to one's diet, would deny that the cookie's appearance is *a consideration in favor of* eating it: "it is no reason at all!" Less metaphorically, we can

describe the better judgment against which the akratic acts as part of an ethical perspective on which the akratic reasons don't show up as reasons.

If one were to present the appearances to which the akratic reacts before the deliberative tribunal, they would be excluded as irrelevant, as not worth weighing. It follows that the judgment made without taking them into account is not an incomplete judgment. The akratic is not in the same situation as someone who, for instance, only notices some valuable feature of one of her options at the last minute. That person notices, just before she acts, that what she took to be her "all things considered judgment" left out something important. The 'reasons' that the akratic leaves out of her judgment are not ones she could have factored into it—for they are not, by the light of her judgment, 'reasons' at all.

Nonetheless, we are evidently capable of being moved by such 'reasons'—and we understand what is happening when someone else is moved by them. Zafar's response to seeing people who look like him is eminently intelligible—not only to us, but also (eventually) to himself. He ultimately does not find his own response absurd or unintelligible, because the point of view from which he spies that ("bad") reason is his own point of view on value. Likewise, it is a point of view the reader can adopt, in imagining what it would be like to have lived Zafar's life. In saying that it makes sense to us that Zafar would react in this way, we are not merely saying that we could predict that someone would have this reaction. We are saying that the reaction has the kind of transparent logic that rational responses have. As Davidson pointed out, when we are able to see "what the agent saw, or thought he saw, in his action," (1963, p.685) we take ourselves to have identified a reason.

III. Which Reason is the Akratic One?

Even once we acknowledge that the akratic's reasons are real enough to have moved her to act but not of a type to be included in the judgment that she ought to act otherwise, there remains some difficulty about the status of that judgment as 'better.' On the standard account, the akratic's better judgment earns the designation 'better' by including, (i.e. having accounted for) the 'weaker' reason on which he acts. On this picture, the authority of the better judgment comes from the fact that it is the akratic's net judgment, reflecting the balance of reasons in his rational bank account. But I have rejected the possibility of an akratic's being motivated by a reason of this kind, any more than someone can akratically choose \$100 over \$200. I want to bring out a cost of accepting my conclusion that an acknowledgedly weaker reason cannot be the reason upon which the akratic acts. On my account, a question arises as to how we identify a response to *one* reason as rational, while counting a response to a reason visible from a different evaluative perspective as irrational. If the 'better judgment's authority does not lie in its inclusiveness, wherein does it lie?

We might try to answer this question by shifting our discussion from talk of 'rationality' to the question of which option represents someone's true or authentic self. Instead of asking which option is backed by more inclusive reasoning, Frankfurt sought to analyze

akrasia in terms of which of the two reasons the akratic *identifies with*⁸³. Frankfurt introduced a higher order act of the will to adjudicate such conflicts: if I act from the part of me I identify with, I act enkratically; action in accordance with the rejected part is akrasia. This suggests that Frankfurt likewise rejects Davidson's conception of akratic conflict, construing it as intrinsic, rather than as resolvable by deliberation. Though Frankfurt shares my sense of the problem, I have already said in chapter three why I cannot accept his solution: he underestimates the depth of these kinds of conflicts. One cannot step back from and adjudicate intrinsic conflicts, whether such adjudication take the form of identification or deliberation. If the conflict between the akratic's two sets of reasons is indeed an intrinsic one, there is no perspective or point of view that encompasses both. I have argue that Frankfurtian identification can't resolve intrinsic conflicts. It follows a fortiori that identification can't resolve that species of intrinsic conflicts which are akratic.

It might help, at this point, to remind ourselves that Davidson's theory was motivated by the aim of avoiding characterizing akrasia as essentially involving a battle between 'reason' and 'passion.' Davidson eschewed any substantive restriction on the kinds of reasons the akratic acts on or against, in favor of a more formal approach in which akrasia is understood as acting on whatever reason one has judged to be outweighed by some other, stronger consideration. The incoherence of the Davidsonian description and the failure of the Frankfurtian solution may incline us to return to the pre-Davidsonian picture of akrasia. On such a picture, the akratic acts on one *kind* of reason (a moral one, or perhaps a prudential one) and against another kind (pleasure, short-term profit, etc.) Davidson's objection to such an account is that it is open to counterexamples like this one:

As a first positive step in dealing with the problem of incontinence, I propose to divorce that problem entirely from the moralist's concern that our sense of the conventionally right may be lulled, dulled, or duped by a lively pleasure. I have just relaxed in bed after a hard day when it occurs to me that I have not brushed my teeth. Concern for my health bids me rise and brush; sensual indulgence suggests I forget my teeth for once. I weigh the alternatives in the light of the reasons: on the one hand, my teeth are strong, and at my age decay is slow. It won't matter much if I don't brush them. On the other hand, if I get up, it will spoil my calm and may result in a bad night's sleep. Everything considered I judge I would do better to stay in bed. Yet my feeling that I ought to brush my teeth is too strong for me: wearily I leave my bed and brush my teeth. My act is clearly intentional, although against my better judgment, and so is incontinent." (1980, p.30)

The story of the toothbrusher is supposed to present us with a reversal of the typical akratic scenario: the 'rational' choice would have been to favor sensual indulgence over prudence, so it cannot be that prudence is always the voice of reason. I am inclined to agree with Davidson that such examples refute the old, 'passion vs. reason' conception of akrasia. Notice, however, the different character of the two 'voices'. Whereas sensual indulgence engages in levelheaded reasoning about the rates of tooth decay at his age, concern for his health motivates him in the form of a strong feeling: "my feeling that I ought to brush my teeth is too strong for me." Davidson seems right to present the

⁸³ For a discussion of how Frankfurt's views on identification/externalization relate to his views on akrasia, see Schroeder and Arpaly (1999), p.374-5.

authority of the ‘better judgment’ as connected to its place in a deliberative process of weighing reasons. But I want to insist that the lack of authority of the other motive comes not because it has been subsumed into, but because it has been excluded from, such a process. Considerations of dental health figure both in the akratic’s conclusion that he should stay in bed and as motives for his akratically getting out of bed—but in very different ways. In order to see this, let us begin by reflecting more broadly on the limited role of deliberation in agency.

Our actions are not always preceded by deliberation as to what we should do. When we act without deliberating, this is not typically because we are indifferent as to whether or not we are acting well, but rather because we take ourselves to have a pre-deliberative answer to the question “what ought I to do?” Often—even typically—someone does not need to deliberate in order to know what to do. Life’s problems are not always difficult to solve. If someone is hungry, she eats. If someone is tired, she sleeps. If she sees her child fall and hurt himself, she runs over to comfort him. She does these all these things for reasons, but she does not need to arrive at these reasons by a process of reasoning. The reasons she acts on in these cases are available to her as soon as she apprehends what is going on. She does not need to take time to step back and reflect on whether things really are as they seem to her; she does not weigh one thing against another. She simply acts. We would not survive very long if we lacked this capacity to respond to reasons without a second thought.

We deliberate because there is some special call to reflect on our situation. This may be because we have not arrived at any predeliberative answer to the question “what ought I to do?” I might be torn between conflicting reasons: my child has fallen, but I’m trying to teach him to be independent; I’m hungry, but I have dinner reservations, etc. Were I not to deliberate, I would have insufficient motivation for either action: I would be torn between two things, both of which I am somewhat inclined to do, and neither of which I am more inclined to do. Perhaps there are even cases in which I’m not inclined to do anything, and must use my faculty of deliberation to introduce reasons into the picture. These are, then, cases in which I deliberate in order to have any answer at all to the question “what ought I to do?” These cases cannot, I contend, take an akratic turn—they are not pre-kratic. If my predeliberative sense of the reasons is null or even, then I am bound, insofar as I act at all, and insofar as the predeliberative situation does not change⁸⁴, to act on the basis of my deliberative conclusion. In this kind of case, my desire to, e.g., comfort my child, is not strong enough, considered independently of the weight I assign this value in my reasoning, to motivate me to act. For that desire, ex hypothesi, did not present itself as an answer to the question “what should I do?” Unlike Davidson’s toothbrusher, I did not have a “strong feeling” as to what I ought to do. I felt inclined to comfort him, and I also felt inclined not to comfort him. If I conclude that I ought not comfort him, then I won’t. Moreover, I will refrain in what Aristotle called a temperate rather than a continent way.

⁸⁴ Of course it might happen that, as I reason, my desire for the one option grows. The point is: I cannot be akratic (or enkratic) unless that desire reaches the strength sufficient to motivate me.

Akratic actions reveal to us that we sometimes deliberate *even when we have a predeliberative sense as to what we ought to do*. A pre-kratic is strongly inclined to some option, for instance, running to help her child. She feels that she should comfort him. But she has an innate suspicion or doubt about this response. She does not trust the answer she predeliberatively supplies herself. She checks her initial inclination to rush over, and reflects on the reasons: he hasn't really hurt himself very much, and though he will stop crying sooner if she comforts him, on balance he will be better off if she lets him see that he can handle this situation on his own. Likewise, the toothbrusher checks his initial inclination to get out of bed, and reflects on whether or not the thing that it *seems* to him he has reason to do—brush his teeth—really is what he has reason to do.

Davidson's description of the akratic gives an inadequate depiction of this initial moment of self-rejection. He presents the akratic as though he were, even at an appetitive level, undecided between the comforts of bed and the attractions of clean teeth: "it occurs to me that I have not brushed my teeth. Concern for my health bids me rise and brush; sensual indulgence suggests I forget my teeth for once. I weigh the alternatives in the light of the reasons..." Akrasia is never so dispassionate. It cannot be that it merely *occurs* to him that he has not brushed his teeth. The akratic must feel a strong independent motivation to act akratically. He is *sufficiently motivated* to brush his teeth, more so than to stay in bed. Brushing his teeth *seems* much more appealing to him. For this inclination was strong enough to drive him to intentionally brush his teeth in spite of his judgment that he should act otherwise.

Anyone who deliberates *despite* the fact that his predeliberative sense of the reasons gives him an answer as to what he ought to do repudiates a part of himself in those deliberations. In deliberatively opening up the question of what to do, he rejects the answer he has already given. He must reject his first answer in order to re-ask the question, and this is why he cannot take that answer into account in the ensuing deliberations. The fact that it seems to one (predeliberatively) as though one ought to brush one's teeth or comfort one's child is, motivationally speaking, a sufficient reason for brushing one's teeth or comforting one's child. The vast majority of tooth brushings and child comfortings happen for this simple reason: it seems best to do so. Davidson's toothbrusher is, in this way, just as 'rational' in brushing his teeth as a toothbrusher who doesn't give the matter a second thought. But this fact—the fact that one has given a preliminary answer to the question 'what should I do'—is one that one commits to seeing as irrelevant once one has begun deliberating. Deliberation purports to improve on the results of one's predeliberative responses, and the 'irrationality' of akrasia lies in retreating to the decision made in the primitive way.

The akratic is intrinsically conflicted between these two ways of answering the question of what to do: the critical deliberative mode and the primitive predeliberative mode. The conflict is the product of the self-rejection contained in the pre-kratic's initial refusal to simply act on the basis of her predeliberative seemings. She decides that there is something wrong with her own way of looking at this situation. She mistrusts her own perspective, her own immediate rational responsiveness, in just the way that Zafar mistrusts his feeling of belonging. He is inclined not to let feeling settle the issue, but to

assess the situation in a detached and rational way, as though he were saying to himself, ‘of course I do not really belong here.’ The agent splits from herself the moment she opens the deliberative question in the face of her predeliberative answer. This split represents the common ground between the akratic and the enkratic. Though the enkratic and akratic diverge on the question of whether they act in accordance with deliberation or predeliberative seemings, they both experience an intrinsic conflict between their continued sensitivity to the fact that something seems predeliberatively to be what they ought to do, and their deliberatively-induced commitment to what they judge best, all things considered.

The phrase “better judgment” is more apt as a description of the judgment the akratic acts against than “all things considered judgment.” The judgment she acts against doesn’t take all things into account—not even if we restrict the “things” to the set of reasons accessible to the akratic at the moment of action. For there is one important “thing” that such a judgment must leave out: the reason that she would have acted on, had she not taken it upon herself to deliberate. But the judgment does indeed present itself as a “better” response to the question of what to do than another available response. The agent’s better judgment is always better than something else, because it always follows upon a prior answer.

An important species of intrinsic conflicts is, then, generated not through the internal incompatibilities of two sets of values but from the mechanics of the deliberative process itself. This analysis accommodates Davidson’s de-moralization of akrasia, while nonetheless helping us see why many cases of akrasia do fit the ‘reason vs. passion’ mold after all. It is in the area of sensible indulgence that we tend to have strong predeliberative responses—and it is also in this area, for that very reason, that we harbor suspicions of being led astray. Our basic desires for food, sleep and sex are, amongst our responses, some of the least informed by reasoning. They are not, in large part, the way they are because of the thinking we have done. We often see these desires *not* as responses to reasons but instead demote them to the status of drives or instincts—even when, from the point of view of those same ‘drives,’ we see ourselves as having a *reason* to, e.g., eat that cookie. Notice that we can make the same kind of demoting move in non-sensibility akratic cases. I can cast my maternal love (as in the fallen child case) as softness, my moral hesitation as inhibition, my religious fervor as prejudice, my well-inculcated habit to brush my teeth as a neurotic attachment to routine. We have all sorts of predeliberative responses to our environment, and we can harbor suspicion with reference to any of them. But we are very likely to suspect those that have been least informed by rational educational processes—or those informed by educational processes that, in retrospect, seem suspect to us.

In these cases, we do not take the rational appearances at face value. The pre-kratic rejects her own point of view; she is a skeptic with respect to the ethical appearances, and thus with respect to herself. She doesn’t trust herself to respond correctly. This lack of trust may or may not be well-grounded. Jonathan Bennett⁸⁵ has drawn our attention to

⁸⁵ See also Arpaly, Bennett, Jones, MacIntyre; all of whom have developed the line of thought found in Bennett.

cases in which someone acts well by acting akratically—so called ‘rational akrasia.’ It is, in the end, not puzzling that someone might have better immediate responses than reflective ones. This is what seems to happen to Huck Finn, who—at least on Bennett’s reading of him—has a head filled with bad moral views, but a heart that is (roughly) in the right place. So he feels strongly inclined to protect runaway slave Jim, but checks that response and reasons that he has a duty to turn him in. He is akratic because he acts, as I would put it, from immediate rationality rather than his mediated, deliberative rationality; but the reasons he acts from are better than those he acts against. They are not better deliberated, but they are better reflections of the moral truth. We will only take the pervasiveness of akrasia as a problem insofar as we think that, by and large, our deliberative processing offers us a better way of solving the problem of what we ought to do than our immediate responsiveness—which is to say, that the Huck Finn cases are outliers. As they do, indeed, seem to be.

IV. Akrasia and Aspiration

Pre-kratic intrinsic conflicts can be conflicts between love and hate, or reason and passion, or habit and morality. Any set of motives can supply the framework for an pre-kratic choice, so long as the one side functions as the agent’s pre-deliberative, the other her postdeliberative⁸⁶ ethical perspective. She cannot deliberate her way out of her prekratic conflict, nor would she be inclined to try. For insofar as she deliberates, she has already decided in favor of one of the two options. Akrasia reveals the fact that an intrinsic conflict may be generated by the very act of deliberation itself.

Pre-kratics are, like most intrinsically conflicted agents, *oriented* toward one of their two perspectives and away from the other. Like aspirants, pre-kratics are not indifferent or confused or unsure as to which way they should go—they are not frozen. But akratics differ from aspirants in that akratics do not feel the need to acquire a better grasp of the value toward which they are oriented. There is, after all, a certain felicity in Davidson’s description of the akratic’s better judgment as her “all things considered” judgment. The phrase “all things considered” connotes a finality that captures the agent’s sense that she knows enough. Davidson’s toothbrusher takes himself to have arrived at an exhaustive understanding of the value of going straight to sleep. The dieter thinks that she fully grasps why she should refrain rather than indulging. The ‘helicopter mom’ takes herself to have a perfectly clear understanding of why she shouldn’t rush to comfort her child. Pre-kratics often express their state in terms of knowledge, “I *know* I shouldn’t...”; and

⁸⁶ A word is due here to the phenomenon of ‘double akrasia’ discussed by Mele (1995) and Jackson (1984). Mele (p.60) describes a boy who yields, against his better judgment that such an action is morally wrong, to his peers pressuring him to go along with a plan to break into an empty house— then akratically backs out at the last minute from fear. Jackson (p.4) discusses a pregnant Catholic who akratically decides to have an abortion, and then akratically fails to go through with it when the time comes. These examples do not undermine the thesis that there is a difference in kind between the akratic’s two motives (or reasons). Rather, they illustrate how short-lived the akratic conflict is. Whether some reason appears in one’s pre- or post-deliberative framework is a question relative to a particular decision context. Having previously, akratically decided to ϕ , and faced with the impulse not to ϕ , these akratics form a reflective judgment that they ought to ϕ rather than yield to their subsequent impulse not to ϕ .

akratics say “I knew full well...” They have a strong sense that their problem, in the situation at hand, does not stem from the fact that they still have something to learn.

They needn’t deny that, more generally, they could stand to appreciate the enkritic value more—some pre-kratics are also aspirants—but they think they appreciate it enough to know that they ought, in the circumstances at hand, to choose the one option rather than the other. Perhaps it is worth dwelling on the fact that some, but *not all*, pre-kratics are aspirants. Though we may be irritated by the fact that we incline against our better judgment, we do not always aspire to attain the condition of the temperate person who remains untempted by chocolate or her child’s tears. Someone may fail to so aspire because she prefers to have the visceral reactions she currently has, seeing in them advantages that outweigh their akratic downside. Or, alternatively, someone might prefer not to have those reactions—“if only there were a button to push!”—but think that a different reaction is not in the cards for her. I may believe, and I may be right to believe, that I am just never going to be the person who jumps happily out of bed every morning, eager to go for a run. In the face of the impossibility of the decisive victory that would constitute reshaping one’s reasons-responses, many of us acquiesce in the project of trying to win as many individual battles as we can. In these cases, we aim only as high as enkrateia.

But suppose someone *does* aspire to become an exercise-lover. She lies in bed, aware that her target-condition is one in which she would be feeling pleasure, rather than dread, at the prospect of rising early into the cold of the dawn. When she looks forwards to the distant future, comparing herself to the person she would like to be, she judges that she doesn’t now have a sufficient grasp of the value of exercise. Nonetheless, she thinks she should get out of bed, despite her dread. She thinks that, given the reasons she *can* grasp for and against exercising, she should go for a run. In order to be akratic, she must take herself to have *enough* of a grasp of the value of her target to judge that it is better to run rather than sleep. Such ignorance as she may also take herself to have as to the true or full value of exercise is here irrelevant. For practical purposes, she knows how important it is to rise.

Why is it that aspirational intrinsic conflicts are accompanied by a sense of defect in relation to the target reason, whereas pre-kratic ones are not? I think the issue is simply one of temporal scale. Akrasia and enkrateia (continence) are aspiration writ small enough that questions of what I hope to learn become irrelevant. The proleptic rationality characteristic of aspiration is directed at the reasons that I take myself to be in the process of acquiring. But no such process can be encompassed within the time limits of the pre-kratic’s intrinsic conflict: for it can only last until she acts. Recall that pre-kratic conflict is made possible by the fact that we can come to an answer to the question, “what should I do?” in two ways: immediately, and reflectively. Once the agent acts, there is no longer a divide between her pre-deliberative and post-deliberative perspectives—until she re-introduces one, when the next temptation rolls around. The akratic is, therefore, different from the bitter wife, who has a *standing* intrinsic conflict that shapes the way she lives, feels and thinks.

I argued, above, that aspiration is the only distinctively agential way to resolve an intrinsic conflict. An akratic's intrinsic conflicts do not last long enough to be amenable to an aspirational solution. We might say they are not agentially resolvable, though their short duration makes this irresolvability less problematic. Or we might say that the pre-kratic does indeed resolve them by acting, but then we will have to grant that she resolves the conflict equally whether she acts akratically or enkratically. I am not sure that it matters which of these ways we look at it—they bring out the fact that the resolution of pre-kratic intrinsic conflict is of a distinctive kind, where this distinctiveness is in turn attributable to the fact that the pre-kratic's target reasons are non-proleptic.

We can now articulate the connection between proleptic rationality and intrinsic conflict. Anyone who has a proleptic reason straddles two value-perspectives simultaneously: the 'old' perspective from which she cannot apprehend any reason to, e.g., listen to music, and the 'new' perspective from which she is moved by the intrinsic value of music. She simultaneously values music and does not value it. The aspirant is not the same as someone who we might comfortably locate at a point along the spectrum from love of music to total indifference. Those who have some relatively weak interest in music might be happy to remain where they stand. They do not feel the defect in their own valuation, and therefore they do not simultaneously inhabit multiple value-perspectives. The aspirant, by contrast, is in motion. Her conflict is one that can be depicted only over a stretch of time, because nothing can move in an instant. Hence the proleptic rationality that characterizes aspiration is confined to *diachronic* intrinsic conflict. A proleptic reasoner is someone who is on the road to a new way of thinking.

We have described the proleptic reason as having 'two faces,' and these two faces really represent the two different rational points of view that the agent moves between. Insofar as she still inhabits the point of view from which music doesn't have intrinsic value, she cannot fully inhabit the point of view from which it does. To the degree that her old self is still present, to that degree is her new one yet absent. Her music appreciation is merely proleptic, in that she can see that she does not respond to those reasons in the way that she's trying to. Likewise, the bitter wife who is trying to be more loving and less hateful sees that the actions and reactions in which her love is manifest are defective versions of the ones she would like to have.

At the end of chapter 1, we discussed the climactic moment in which an aspirant must face the fact that pursuing her aspirational aim may come at a substantial cost. At these moments, she is called upon to make use of whatever grasp of the value she has already acquired, so as to weigh it against the value she is being asked to sacrifice. Such conflicts—as between one's Zionism and one's duties to one's family, or childbearing and career, or college education and living debt free—are extrinsic. Someone could, in principle, step back and survey both values; the agent's difficulty in doing so stems from the fact that she grasps (at least) one of them imperfectly. She cannot fully adopt the deliberative perspective of one who appreciates the value to which she is still in an aspirational relation. Because she is intrinsically conflicted over her own valuation of, e.g., motherhood, she finds herself ill-situated to evaluate it. And this problem—the fact that she cannot get the value of motherhood fully into view—is one that she cannot

resolve by deliberating. Because she is forced to make a choice, she may have to step back, reflect, and do her best to assess the value of motherhood. Such reflection cannot, however, contribute to the aspirational resolution of her intrinsic conflict.

Proleptic reasoners are like akratics in experiencing a conflict that deliberation cannot resolve, but differ in that aspirational conflicts are not based in the mechanics of deliberation itself. Rather, aspirational conflict is grounded in the fact that one is in the process of moving between two perspectives on value. Aspirants inhabit conflicting perspectives not only at a moment, but as part of a larger project of self-change in which they are engaged. In the third and final section of this book (chs.5-6), I will examine the ethical implications of the fact that such projects are available to us. There it will emerge that the fact that intrinsic conflict *needn't* be confined to the timescale of akrasia—i.e., the fact that intrinsic conflicts can involve proleptic reasons—offers the theorist of moral responsibility a novel avenue for defending the coherence of the concept of self-creation.

Chapter 5: The Problem of Self-Creation

In the first section of this book (chs.1-2), we considered aspiration from the vantage point of decision theory: how it could be rational for someone to pursue a project that is not the optimal way of satisfying her current set of preferences? Our answer was that the rationality of aspiration is not capturable within the synchronic framework of decision. The aspirant's reason is proleptic, with a proximate face that speaks to the person she is now, and a distal face fully visible only to the person she will become. Proleptic rationality is essentially extended in time, being a form of value-learning.

We then turned (chs.3-4) to the moral psychology of aspiration, attempting to give an account of what proleptic rationality feels like. The aspirant is trying to see the world through another person's eyes, namely, through the eyes of the person who has the value she aspires to acquire. Her condition cannot be captured in the Frankfurtian framework: she neither 'identifies with' nor is 'alienated from' the evaluative perspective she imperfectly inhabits. Rather, her condition is one of intrinsic conflict between the evaluative perspective she seeks to acquire, and the one she seeks to depart from. Further, we noted that intrinsic conflict is broader than aspiration: the akratic is also intrinsically conflicted.

In this final section of the book, we will examine the implications of the phenomenon of aspiration for the theory of autonomy. The aspirant is not simply someone who acquires a new value, but is also the one responsible for orchestrating the value-acquisition. She gets herself into a different value-condition from the one she is in. From this vantage point, aspiration emerges as a kind of *work*. The work is visible in her struggles to sustain interest in the hobby or relationship or career or religion or aesthetic experience that will later become second nature; in her repeated attempts to 'get it right,' attempts that must be performed without the benefit of knowing exactly what rightness consists in; most generally, in the fact that she always wants and strives to be farther along than she is. An aspirant's value-transition is her own work, which means that she is a certain kind of cause: a cause of herself.

In saying this, we cannot deny that the aspirant receives help. Major value-acquisitions reflect the influence of one's environment, especially the people in it. Parents, teachers and lovers have transformative effects on the people they parent, teach and love. But they cannot, at least typically, have these effects on someone unless she participates in the process. They assist, rather than substituting for, the activity of the agent herself. My aim in these two chapters is to offer an account of the work an agent does in making herself into a person with new values. I call this work "self-creation."

That label may be sound overblown, given that the person must have been around to do the creating. Nonetheless, she didn't then exist *as* the person she becomes. The

advantage of “self-creation” over “self-change” or “self-revision,” as I might also have called it, is that the former phrase serves as a reminder that we are not interested in minor or superficial ways in which someone might change himself. Self-creation in the sense I’m interested in is not going to be a matter of a person making physical changes to his body, altering his musculature or the color of his hair. Caring about something new can have physical repercussions, but the change one is effecting on oneself is not in the first instance a physical change. Nor is it a psychological one, though it will have psychological repercussions: one’s new form of valuation lends itself to new emotions, feelings of attachment and vulnerability, curiosity, interest, excitement. But these, too, are consequences of something deeper, which is the *ethical* change occasioned by committing oneself to some form of valuation. Our interest is in self-directed value acquisition, which is, first and foremost, a change of a person in the ethical dimension. In this chapter, I propose to explain how the aspirant can be an ethical source of herself.

I begin by explaining why this question is so difficult to answer. I describe two received models for self-directed agency, and argue that neither offers us a framework for understanding self-creation. First, I consider the possibility that one ‘creates oneself’ by an act of the will in which one endorses some way of being, sanctioning it with an evaluative stamp of approval. I call this the ‘self-endorsement model,’ and I argue that because one is in a position to endorse only what one *already* values, self-endorsement cannot represent a way of *acquiring* values. Next, I turn to what I call the ‘self-cultivation model.’ Self-cultivation is the process of working to satisfy anterior normative commitments about what kind of person to become. I adapt a paradox described by Galen Strawson to show that we cannot tell a story of genuine self-creation as a story of self-cultivation. If, on the one hand, the value I cultivate in myself follows rationally from values I already have, then I do not do any creating. For in this case my “new” self was already contained in my old self. If, on the other hand, the new value is rationally unconnected to my earlier values, then its advent in my life cannot be my own doing. In this case, the self I end up with may be new, but it is not the product of my own agency. Rational value-cultivation is not *self-creation*, and non-rational value-cultivation is not *self-creation*.

I analyze the underlying problem as follows. Call the self that does the creating “S1,” and call the self that is created “S2”. For the theorist of self-endorsement, S1 is the endorsing self, S2 the self it endorses; for the theorist of self-cultivation, S1 is the cultivating self, S2 is the self it cultivates. Both theorists depict S1 as normatively prior to S2 in the following sense: they present S1 as the agent’s authoritative self, the self whose agency (via endorsement or commitment) determines how S2 ought to be. I argue that if we attempt to derive the norms governing S2 from the values or commitments already present in S1, we will never be able to describe self-creation. Instead, one must reverse the priority relation between the two selves. In section IV of this chapter, I show that we can do so by conceiving of the later self as normatively prior to the earlier one. The aspirant does not see herself as fashioning, controlling, sanctioning or shaping the self she creates. Instead, she looks up to that self, tries to understand her, endeavors to find a way to her. Thus the process of self-creation is distinctively aspirational.

Skeptics of self-creation, Strawson among them, are wont to cite with approval a passage in which Nietzsche scoffs at the philosophical impulse “to pull oneself up into existence by the hair, out of the swamps of nothingness.” He describes self-creation as a “rape and perversion of logic⁸⁷.” Such skeptics may worry that an account such as mine succumbs to the basic error Nietzsche is describing: in positing the normatively prior self as temporally posterior, I am presenting a teleological account of (a species of) self-directed agency. Is such teleology naturalistically suspect—does it rest on some notion of backwards causation, or causeless effects? I argue that these suspicions are themselves predicated on a conceptual error: a confusion between two kinds of grounding relation. In the case of S1 and S2, we are asking questions about how one set of norms—the norms governing S1—is related to another set of norms—the norms governing S2. We need to distinguish this question about normative grounding from any question as to *causal* grounding. Nothing prevents the theorist of aspiration from offering a traditional (i.e. non-teleological) account of the causal grounding of the genesis of S2 in the representations, desires, etc. of S1. If we want to know whether those very desires and representations succeed or fail, we must assess them with reference to the as-yet non-existent S2.

In aspiration, it is the created self who, through the creator’s imperfect but gradually improving understanding of her, makes intelligible the path the person’s life takes. Aspiration is that form of agency in which one acts upon oneself to create a self with substantively new values. One does this by allowing oneself to be guided by the very self one is bringing into being.

I. Self-Endorsement

A number of philosophers have taken an interest in our power to endorse, or withdraw approval from, some feature of ourselves. Harry Frankfurt distinguishes between two kinds of drug addict: the unwilling addict rejects his own impulse to take a drug, by contrast with the willing addict who approves of his own addiction. Frankfurt (1971) says that the former does not desire as he *desires to desire*; and (in Frankfurt 1988), the latter *identifies* with his desire. Christine Korsgaard might describe that same addict as endorsing (or failing to endorse) the practical identity of being an addict (Korsgaard, ch. 3). Gary Watson covers the same territory by contrasting a person’s evaluative and motivational systems: the former issues judgments as to the value of the desires populating the latter. If approval is withdrawn, as in the case of the unwilling addict, the agent is “estranged from his ...inclinations” (p.210).

Abstracting from their differences, we can group these views together as describing *self-endorsement*, i.e. an activity in which an agent, or some part of her, steps back from, appraises, and attaches a positive or negative evaluation to the aspect of herself she evaluates. I do not doubt that self-endorsement is a real and significant aspect of agency,

⁸⁷ *Beyond Good and Evil*, section 21. Are these skeptics are right to see Nietzsche as an ally? He is, after all, apt to wax eloquent on the power of the strong to create themselves. See the passages cited in Owen and Ridley (2003) in response to the more straightforwardly fatalist reading offered in Leiter (2001).

but I will argue that it cannot serve as a means of self-creation. We do not change who we are by reflectively approving or disapproving of ourselves.

Self-creation, as I understand it, involves the creation or destruction of values. But values are also what we use in forming our endorsements or rejections of some feature of ourselves. Can one use one value to assess another? Consider some possibilities for how this might go. Suppose that I have exactly three values: V_1 , V_2 and V_3 . I wish to evaluate V_1 from the point of view of V_2 , or from the point of view of V_2 and V_3 . I cannot simply ask myself, “is V_1 valuable?” for then I will be putting my finger on the scales. I *do* value V_1 , and will end up evaluating V_1 from a point of view that assumes the value of V_1 . But I might successfully set that value aside by asking myself a hypothetical question: “Should someone who only values V_2 also value V_1 ?”; or, “Should someone who only values V_2 and V_3 also value V_1 ?” In this way, I can investigate the question of whether to value V_1 without presuming its value. But notice that if I answer “no” to both of these questions, this has no implications for what I should do or value: I am not someone who only values V_2 , or who only values V_2 and V_3 . The hypothetical model allows an agent to evaluate one value from the point of view of another, but at the expense of being uninformative of her actual situation. Such hypothetical reasoning does not put the agent in a position to change her value-endorsement.

This observation might lead the theorist of self-endorsement to abandon the project of evaluating a value by way of some value set that does not include it, in favor of evaluating a value by looking at its place in one’s value-system as a whole⁸⁸. One can observe, while still valuing X , that one’s valuation of X does not cohere well with the rest of what one values. Suppose, then, that someone realizes that one of her values—say, her valuation of her appearance—detracts from her ability to value something else—say, equality or independence—to which she is more fundamentally committed; or that given her feminist values she really ought to value a certain form of political activism to which she is currently indifferent. Do we, in this case, have a form of self-endorsement that amounts to self-creation?⁸⁹

⁸⁸ Laura Ekstrom (1993) modifies Frankfurt’s account by adding such considerations of coherence to it. She aims thereby to defend Frankfurt against Watson’s charge that any higher-order desire must itself be endorsed by a yet higher-order desires, generating a regress. On Ekstrom’s picture my desire is “mine” because I “authorize” it in virtue of its occupying a position of coherence amongst my desire set: “The coherent elements “fit with” the other items one accepts and prefers, so that, in acting upon them, one is not conflicted.” Ekstrom’s defense suffers from a systematic ambiguity as to whether what makes a desire mine is the fact that it coheres with the rest of my desires, or the fact that that I take it to cohere with them and therefore authorize it: endorsement and coherence compete for centrality in her account. Even if it worked, however, her defense could not go so far as to offer a theory of self-creation. For Ekstrom identifies a person’s values with those desires that “fit with” one another (“I propose to define coherence with one’s character system...as determined by what it is valuable to pursue within that system.”, p.610) and this means that one cannot so much as articulate the question as to which values, i.e. which (among many) coherent system of desires, one should “authorize.”

⁸⁹ Supposing, for the sake of argument, that it really is rational to become more coherent. For an argument against the existence of demands of this kind, see Kolodny “Why Be Rational” and “Why Be Disposed to be Coherent.”

No: the realization in question stops just short of endorsement. We are to imagine the agent asking herself questions such as: “I know I value X, but *ought* I to value it, given my valuation of Y and Z?” Or, “I know I don’t value W, but *ought* I to value it, given my valuation of X, Y and Z?” If she concludes that she ought not value X, or that she ought to value W, this does not constitute an endorsement of those values, but rather a change in her judgment as to what she ought to value. Such an agent comes to see that she is not as she ought to be. The project of assessing oneself for coherence presupposes, at the point of asking oneself these questions, the possibility of separating the question of what values one has from the question of what values one should have. That same separation must, then, reappear when the agent arrives at an answer. One can only conclude that one shouldn’t have a certain value *that one continues to have*.

The judgment that she would be more coherent if she valued X cannot itself constitute an endorsement: a new endorsement would entail a change in what she values, whereas the assessment in question works by holding the question of her actual values fixed. Nor can we say that the formation of the judgment that I ought to value X amounts to the creation (or endorsement) of a second-order value. If there is such a thing as valuing a value, this value cannot be what is produced by my coherentist self-policing. Rather it must, like my first-order valuing of X, figure among the objects I police. When I ask myself what I must value in order for my values to be coherent, I must step back from my higher-order values as well as my lower order ones, since I am asking myself not what I *do* value, nor what I *do* value valuing, but what I ought to value, or ought to value valuing.

It might seem, however, that coherentist self-policing is but one small step away from value change. The realization that it would be rational to gain or lose a value may not itself amount to value-change, but it could constitute an impetus to value-change. In the face of her acknowledgement that she is not the person she ought to be, the agent might be moved to go ahead and (commit to) change herself into that person. In the next section, I will consider whether we can model self-creation using the process of bringing oneself into line with one’s judgments about one’s values. This process includes, but extends beyond, the moment of self-assessment to encompass the agency that flows from and is governed by whatever assessments one makes.

II. Self-Cultivation

In chapter I, I described a form of preference change that I called “self-cultivation,” in which one engenders preferences in oneself because one antecedently appreciates the value of having those preferences. When we cultivate ourselves, we don’t just stand back and assess ourselves, we actually try to go ahead and change ourselves. Thus self-cultivation might seem to be a more promising model for self-creation than mere self-endorsement. Philosophers such as Jean Hampton, Michael Smith, Michael Bratman, Richard Holton, and Joseph Raz have all articulated what can, broadly speaking, be characterized as theories of self-cultivation. For they all describe a process by which a person can act so as to determine her future self.

Hampton describes a “self-authorship” process: “There are many times in our lives when we choose what we will be. For example, when a young girl has the choice of entering

into a harsh regimen of training to become an accomplished figure skater, or else refusing it and enjoying a more normal life with lots of time for play, she is being asked to choose or author whom she will be. When a graduate student decides which field of her discipline she will pursue, or when a person makes a decision about his future religious life, or when someone takes up a hobby--all of these choices are ways of determining one's traits, activities, and skills, and thus ways of shaping one's self---of determining one's self- identity." Hampton contrasts this "authentic" form of self-determination with one in which a person allows herself to be shaped by social or environmental pressures into becoming what people want her to be. "Self-authorship involves.... a decision to develop the traits, interests, and projects that are not only consistent with meeting your objective human needs but that are also ones you want, and not ones that others prefer that you want (and perhaps try to persuade you to want)." Thus she grounds the process of self-authorship in the desires of the authoring or creating self.

Smith imagines self-cultivation as a reflective process, akin to the coherentist self-policing described above, by which one moves oneself toward a condition of greater consistency amongst one's desires: "Systematic reasoning creates new underived desires and destroys old. Since each such change seems rationally required, the new desiderative profile will seem not just different from the old, but better; more rational. Indeed, it will seem better and more rational in exactly the same way, and for the same reasons, that our new corresponding evaluative beliefs will seem better and more rational than our old ones." (p.116) Bratman develops Frankfurt's higher order model in a diachronic direction. He presents the adoption of higher order self-governing policies as a solution to the problem he calls "underdetermination by value judgment," i.e., the situation in which one is confronted with a choice amongst what are, from the point of view of one's antecedent desires, a multiplicity of equally valuable pursuits. When there are many good things you could do, committing yourself to one can settle which you subsequently ought to do. Holton defends the rationality of "rational non-reconsideration," (p.3) by which a person adheres to earlier resolutions as to what she should now do, and does not open the matter for deliberation. Raz, whose view I will describe in more detail below, describes the agent at the output of the process of cultivation as doing something akin to keeping promises made by her past self. In all of these cases, an agent makes and subsequently lives under the normative guidance of some kind of choice as to how she should live.

We certainly do cultivate our physical appearance, friendships, desires, hobbies, traits, skills, traits, activities, and careers. But the difficult case is, once again, the cultivation of the state of valuation from which the agent will select *this* physical appearance, *that* friend, rejecting *this* desire in favor of *that* one. Can someone cultivate value in herself? In the course of an attack on the possibility of moral responsibility, Galen Strawson provides the materials for an argument that value cultivation is subject to the following dilemma⁹⁰.

⁹⁰ Strawson's ultimate target is responsibility for action: he wants to show that because we cannot create our values, we cannot be responsible for anything we do from those values. I discuss the major premise—

The new values, acquisition of which constitutes my act of self-creation, must either be continuous or discontinuous with the ones I already have. If they are continuous, my new values are entailed by my old ones. In this case, I don't really change. If they are discontinuous, the new values contradict or come at a tangent to my old values. In this case, they arise accidentally or through external influence, rather than through my agency. I change, but I do not change myself.

Let us see how this dilemma plays out in a schematic case. Suppose that I value E, and I come to value M because I see that M is the means to E (or more generally, that valuing E rationally requires me to value M). Will this count as self-creation? No: instrumental reasoning works out the consequences of the value condition I already have. It does not discover ends, but only means. But what if, prior to realizing the relationship between E and M, I actively disvalued M? Overcoming my hatred of M might then constitute a substantive change—a change in my ends. But in that case, the mere discovery that valuing E rationally required me to value M is insufficient to explain the change. For why didn't I abandon E instead of embracing M? Perhaps I have *other* values that dictate this decision, or perhaps I simply valued E more than I disvalued M. In either case, we are once again back on the first horn: this isn't a change in value but merely a working out of the consequences of values I already have. If I neither valued E more than disvaluing M nor had other values dictating the overall importance of E, then we have not yet explained my rationale for adopting M. I may have done so randomly, or on a whim, but if my new self is the product of such arbitrariness, it is not something I made. Strawson makes this observation when considering a libertarian defense of free will. He asks, "How can the occurrence of partly random or indeterministic events contribute in any way to one's being truly morally responsible either for one's actions or for one's character?"(p.18) This point applies equally to cases in which one's efforts of self-change are made for no reason, or arbitrarily. As Strawson points out, someone who succeeds as a result of such an event is "merely lucky." He has ended up with a good self that he did not create.

At each point in the story of E and M, we face the same choice: if the value I engender in myself follows rationally from values I already have, then I do not do any creating. For in this case my "new" self was already contained in my old self. But if the new feature is rationally unconnected to my earlier values, its advent in my life cannot be my own doing. In this case, the self I end up with may be new, but it is not the product of my own agency.

Why does the fact that the value that comes to exist at t2 was derivable from the agent's values at t1 disqualify the case from counting as self-creation? We should allow that the derivation might have required time and effort, so that the agent at t2 differs substantially

that action-responsibility derives from character-responsibility—below, p. 147. Since he does not always clearly separate the arguments for these two claims, and since I am not discussing moral responsibility, what I offer here is not a direct paraphrase of any passage of his paper. Nonetheless, both style and the substance of my presentation of the dilemma is heavily indebted to Strawson's way of framing (what he takes to be) an objection to both self-creation and moral responsibility.

from the agent at t1. Coming to value something that is instrumentally conducive to ends one had all along might occasion observably significant changes in acting, thinking and feeling. But if we allow that such an agent develops her new self by unraveling the implications of the materials present in her old one, we will find that we have simply pushed the question about self-creation back to an earlier stage: how did she acquire those materials? And so the problem is not that we must deny that the agent at t2 is self-created; the problem is that the process seems to have gotten going before t1. Her valuation of the means is a *new expression* of a value that she acquired at whatever time (before t1) she came to value that end. We will not have an account of how her ethical self came into being until we have a story of *that* genesis. But if we try to tell one, Strawson's dilemma recurs.

We encounter the same regress if we press, on Bratman's behalf, the question of why a value choice underdetermined by one's previous commitments cannot constitute self-creation. Suppose Sartre's soldier was not antecedently more committed either to his mother or his country. Forced to choose, he decides (at t1) to devote his life to mother over country. Does it really follow from the fact that it would have been equally rational (at t1) to opt to fight that the filially pious self emerging at t2 is arbitrary in relation to his self at t1? Bratman might point out that this would only follow if at t1 he cared little or nothing for his mother. Suppose to the contrary, that the soldier at t1 already cared deeply for both mother and country. In this case the value must already have been created sometime before t1. The decision made at t1 is then not value creation, but one (of at least two) possible developments of his antecedent state of motivation. If we want to know how *that* state got created, we will have to look back further into his past. And then Strawson re-poses his dilemma.

Anytime we tie some new value to an act of self-cultivation, we must admit the arbitrariness of the choice, or push the source of the cultivation back a further step. And this regress is a vicious one: as we retreat backwards through a person's selves, we encounter selves that are less and less, and eventually not at all, in a position to do any creating. And shortly thereafter, of course, we run out of self altogether.

Must my later self either come at a tangent to my earlier self, or be a rational consequence of it? If the two horns of Strawson's dilemma were exhaustive of the possibilities, self-creation would be—as Strawson thinks it is—a chimera. I will argue that they are not, but I want to note that Strawson's conception of the space of possibilities is widely—though not always explicitly—shared by theorists of the self. Consider, for instance, Robert Noggle's description of “the basic picture of the self” that is “quite common among philosophers.” (fn. 26) Such a self “evolves according to its own internal logic—its own contents determine whether and how it is to change in response to new information, internal conflicts and changing condition....when psychological changes happen in this way, it seems correct to say that the new configuration is an authentic continuation of the previous configuration. On the other hand, a psychological change—especially a change in the core attitudes—that does not occur in this way produces a new configuration that is not an authentic continuation of

the previous one...If the changes are radical enough, it might be proper to speak of the destruction of one self and its replacement by a new one.” (pp.100-101)

As examples of changes in the second category, Noggle offers “sudden organic trauma” and “nefarious brain surgery.” He contrasts these with changes that are “internally motivated in such a way that they seem to be *intelligible reflections* of the contents of the core attitudes (p.100, emphasis in original).” Noggle’s division recapitulates Strawson’s dilemma: he sees changes in a self as either rational extrapolations of previous conditions (“according to its own inner logic”) or traumatic intrusions from without. Noggle’s discussion of self-creation showcases how deeply we have been gripped by Strawson’s dilemma. The two selves must be rationally connected, or we cannot see the second as the work of the first; but this same continuity prevents novel values from arising.

III. Neurath’s Boat

At this point, we may be tempted to conclude that self-creation is, indeed, impossible. Why should we demand that we be able to do more than revise our selves in the light of our current values, becoming more consistent versions of the people we are? We may be inclined to heed Nietzsche’s advice to stop trying to pull ourselves “up into existence by the hair, out of the swamps of nothingness.” In order to pull ourselves up by the hair we would have to be able to detach ourselves from our bodies. And this may seem to be precisely the logical problem at the heart of the idea of self-creation: we cannot step outside ourselves. It has become a commonplace to describe the developing self using the image of Neurath’s boat, which must be rebuilt as it sails⁹¹. It is impossible to fashion oneself as though one were not already some person with desires and judgments; moreover, with desires and judgments that speak to the question of how one ought to fashion oneself. One can, at most, refashion from within. I want to take a moment to motivate the claim, which may at this point seem questionable, that there is indeed a *phenomenon* of self-creation for which we should try to account. Let me begin by pointing out that Neurath’s boat is not Theseus’ ship.

The fact that Neurath’s boat never docks is indeed an elegant way of expressing the unsheddability of character. One cannot ‘operate’ on one’s self from the outside, as a doctor might operate on his own anaesthetized leg. One cannot set aside one’s way of looking at the world and somehow examine things from the outside of one’s self; as we saw in our discussion of the power of reflection in chapter 3, reflection does not afford a person the opportunity to ‘step back’ from her values. But it does not follow from the fact that reflection has limited power that the self must stay true to its initial form, or else be traumatically violated from without. Neurath’s boat isn’t Theseus’ ship, whose job is to stay the same through all its many changes. On Theseus’ ship, a rotting plank must be replaced with a plank of a similar shape and size. On Neurath’s boat, the only requirement is that it must be replaced by sailors on board the boat. Neurath’s rowboat might be gradually rebuilt into a trireme, or, for that matter, an airplane, without the builder ever setting foot onto dry land. The fact that there is no vantage point one can

⁹¹ See Noggle fn. 26, McDowell pp.36-7.

simply adopt outside one's character doesn't entail that one couldn't *arrive* at the vantage point that is outside one's current character, by working toward that condition.

It is a fact of life that we can do more than maintain ourselves: we can build ourselves up into something better. We become spectators, we take classes, we do exercises. We find mentors to emulate or fellow travelers with whom to commiserate—and compete. We do the same thing over and over and over again until we get it right, without knowing in advance what 'right' is. We do work we don't always enjoy, and we pretend—even to ourselves—that we enjoy it. We leave ourselves open to certain kinds of experiences and closed to others, knowingly risking disappointment and disillusionment down the line. We alert ourselves to and steel ourselves against temptations to abandon course in favor of a more readily available and more immediately intelligible form of value. Candy, television, alcohol, a nap, video games, internet surfing—pick your poison, it's waiting in the wings. We struggle against implicit or explicit messages, from individuals or groups of individuals, to the effect that this kind of value is "not for you." Often these struggles are heightened by the fact that we have internalized the judgments in question. The work that we are engaged in is the work of bringing something into view. But because what we are bringing into view is something practical—a value—the work is a matter of acting, and feeling, as well as thinking.

This kind of work involves both moving toward and moving away from a perspective on value. When engaged in it, we are not only gaining something, we are also often losing something. In some cases, the value perspective that we are losing can be characterized relatively thinly as the non-appreciation of something. But sometimes an agent can identify strongly with their pre-aspirational valuation condition. Sometimes people view their own indifference to art or music as connected with being unpretentious and straightforward. Adults who describe themselves as 'childfree by choice' view their condition not in terms of the absence of children, but the presence of the freedom to dedicate themselves to work or friendship or romantic love or travel. The bitter wife is not merely lacking in love but actively *hates* her husband. Perhaps she feels, given the way he treats her, that it is only by withholding her affection from him that she can respect herself. Alcibiades sees Socratism as the loss of the only life he knows how to lead, the life of honor. People in these conditions have a strong barrier to changing in respect of the value in terms of which they have identified themselves. They feel, possibly correctly, that there are many good things they would be giving up by acquiring the value in question; and that, because the value is one with which they have (at most) only aspirational contact, they would be giving up these known goods for an unknown good.

It is, on the one hand, a testament to the flexibility of the human soul that even people who build a sense of identity around not valuing something (music, children, one's husband) can sometimes bring themselves to adopt the opposed value-perspective. People fiercely attached to independence can become wonderful parents. And couples who could not get their future into view without placing children at the center of it can struggle past the deep sadness of infertility to embrace the freedom of childfree life. It is, on the other hand, a testament to the difficulty of perspectival change that sometimes

such people find themselves unable to see things differently, a fact which infuses the rest of their lives with a sense of loss. It seems possible to work to come to see a value one didn't see before, as well as to divest oneself of the value perspective one currently has: this is the process I have been calling aspiration.

On an aspirational account, self-creation is agent-driven learning in the domain of value. The aspirant brings herself to a different view as to what matters in life, and comes to appreciate what she once did not. How does this contrast with self-cultivation? Consider an example with which Jean Hampton illustrates her "self-authorship" conception of aspiration:

"in the spring of 1991, American newspapers recounted the story of an investment banker who, as a teenager, wanted to be a clown. His parents strongly discouraged it, regarding it as inappropriate for someone of his background and abilities, so he went to MIT and got a job working in Silicon Valley in computers. Still he was dissatisfied and decided things might go better if he had an MBA. With this degree he got a job on Wall Street making a lot of money in a high-powered investment bank. But one day, he claimed, he woke up realizing that if he kept working on Wall Street, he would end up close to death never having gone to clown school. So he quit his job, and did exactly that. This is a nice story of someone who struggled to author himself, while under pressure to be what people in his social group expected of him... he faced pressure to submit to a social role, to take on preferences, interests, and projects that he did not really want. He experienced understandable relief when he reclaimed himself."

The striking feature of Hampton's Banker/clown, from an aspirational point of view, is how little learning he had to do. His challenge lay in resisting the social pressures and exercising autonomy in the service of doing what he really wanted all along. His self was already, as a teenager, developed in the direction of the core value of his life. He did not have to work to discover what he wanted; he had but to claim it, or later, to "reclaim" the value that was present in him all along. Such a picture focuses our attention away from the developmental period during which one learns to see something in, e.g., clowning. I do not deny that one important way in which autonomy manifests itself is in making choices that align with who you really are, as opposed to who others want you to be. But this is not the manifestation of autonomy, for it presupposes the existence of "who you really are." On the aspirational model it can be true that you are not yet "who you really are;" your wants are themselves a work in progress.

One place to see the difference between self-cultivation and aspiration is by looking to the role allotted to secondary players. In Hampton's example, other people appear in the form of "social pressure" to abandon one's true desires. No doubt the man in question also had friends, family or associates who encouraged and perhaps facilitated his ultimate career change. One needn't, however, make mention of such people in order to tell this story of "self-authorship." He is in an important respect sufficient unto himself, equipped from the start of the story with knowledge of what he wants. This is the arena in which the aspirant needs help from others. Tales of aspirational self-creation will, of course feature the aspirant in a starring role, but they also lean heavily on a host of supporting players: teachers, mentors, (supportive!) parents, schools, advisors. Everyone relies on the care and love of the people around them, but aspirants rely on the people around them

to care about and love the things they themselves are struggling to come to care about and love.

Compare aspirational help to medical help. The patient doesn't know how to cure herself—on this point she lacks knowledge or ability or resources. But there is something she does know: she wants to be cured of the disease. In this kind of case, the agent's practical ignorance or inability is circumscribed by her practical knowledge or ability with respect to her desired end. She knows for certain that she wants a cure. It is not essential for her to acquire the knowledge or ability she lacks: she doesn't need to cure herself. In this sort of case, the person's ignorance/incompetence and her knowledge/competence can be cleanly separated from one another. The aspirant's ignorance, by contrast, runs 'all the way down,' in the sense that she cannot be sure, until she is no longer an aspirant, that she even wants to acquire the relevant value. And insofar as she has not grasped something about the value, she sees it as her job to come to do so. Aware that they have something to learn, aspirants lean on those more securely attached to the value in question, but even their leaning is aspirational, displaying a pattern of ever-lessening pressure. (See above, ch.3 II,3) Like a child learning to ride a bike, the aspirant gradually replaces her helper's support with her own, internal balance. This delicate interplay between one agent and others is one among the features of self-creation that is obscured if we assume the model of self-cultivation. And it is obscured, as I will argue, by the assumption that the created self (the endorsed or cultivated self) depends rationally on the creator self (the endorsing or cultivating self)—as opposed to vice versa.

IV. Grounding Relations

Kit Fine, Paul Audi and number of other metaphysicians have recently drawn attention to the usefulness of invoking a two-place relation of determination that they call "grounding."⁹² We typically express this relation using phrases such as "in virtue of" or "makes it the case that" or by specifying a nontemporal kind of priority, a non-causal use of "because," or a non-spatial type of following. Consider two examples from Plato's *Euthyphro*. When Socrates asserts that something is carried *because* someone is carrying it (10bc), he is, perhaps surprisingly, denying that the fact that A is carrying B is the very same fact as the fact that B is being carried by A. Instead, he takes the former fact to ground the latter. Likewise, when Socrates asks Euthyphro whether the gods love the pious because it is pious, or whether something is pious because the gods love it (10a), he is asking a question about grounding. When Euthyphro assents to the former (10d), he thereby agrees that the piety of some action grounds the gods' love of it.

Both Fine and Audi distinguish a variety of forms that such grounding might take⁹³, and pick out a species of grounding they call "normative grounding." Here is an illustration, adapted from Fine:

⁹² See the entry in the Stanford Encyclopedia of Philosophy (Bliss and Trogon, 2014), and the papers in Correia and Schnieder 2012.

⁹³ Though they disagree as to whether these are different kinds of determination (Fine), or different cases of one underlying kind of determination (Audi).

NG1: the fact that someone's action is immoral obtains in virtue of the fact that he acted with a bad intention.

NG1 says that the more basic normative fact that he intended wrongly underwrites or serves as the normative ground for the wrongness of his action. We can paraphrase NG1 by saying, "what he did was wrong *because* of the intention with which he did it" or "his intention is what made his action wrong"; or "the wrongness of his action follows from that of his intention." These are all ways of expressing a normative grounding relation between one normative claim and another normative claim. This intra-normative grounding relation is the one I will pick out with the expression "normative grounding." Someone might also take an interest in grounding relations between normative claims and non-normative claims, for instance if he wants to understand how (and whether) some form of normativity is itself grounded in underlying causal or metaphysical facts. I will have occasion to discuss such extra-normative grounding claims in VIa below, but for the moment my attention will be on grounding that is normative through and through. Fine proposes that attention to such grounding relations amongst normative claims is the province of ethics in the way that attention to what he calls "natural grounding," i.e. the logic of causal relations, is the province of natural science. In what follows, I hope to offer a concrete illustration of how an ethicist might make use of the framework of grounding to clarify and resolve a paradox.

Note, first, that "normative" is being used in a broad sense. Consider:

NG2: The rules of baseball are normatively prior to any particular game of baseball.

NG3: Her action is rude because it violates club procedure.

NG4: Whether or not the soldier should have fired depends on what his commanding officer instructed him to do.

NG5: Beauty is not what gives an artwork aesthetic value.

These are all claims about the presence (NG2-4) or absence (NG5) of normative grounding relations. NG2, for instance, says the fact that the rules of baseball are such-and-such makes it the case that such-and-such events constitute a game of baseball. In a relation of normative grounding, one fact or state of affairs governs the application of norms (of morality, sports, aesthetics, etiquette, martial order etc.) to another fact or state of affairs. The normative status of what is grounded depends on the normativity inherent in what grounds it.

The language of normative grounding allows us to articulate a particularly important kind of philosophical dispute, namely, a dispute between two theorists, one of whom thinks that A normatively grounds B, the other of whom thinks that B normatively grounds A. As an illustration, we can use the NG1 to frame a debate between a certain kind of Kantian and a certain kind of Aristotelian. The Kantian I'm imagining says, with Fine, that the worth of the intention grounds the worth of the action. On her view, intentions contain a principle on which the agent conceives of himself as acting. The worth of the

intention depends on the universalizability of this principle (or “maxim”); this worth then grounds the worth of the action.

By contrast, one might imagine a (certain kind of) Aristotelian who holds that actions are the proper objects of judgments of goodness and badness, i.e., praise and blame. The goodness or badness of actions grounds the goodness or badness of character, in that a good character is the character that disposes one to perform good actions; character, in turn, grounds the goodness or badness of an intention, in that a good intention is whatever intention characteristically issues from a good character. Among other differences, these two theorists have a dispute about the location or source of moral worth: the one holds that actions have moral worth because (and insofar as) they are associated with certain kinds of intentions, and the other that intentions have moral worth because (and insofar as) they are associated with certain kinds of actions.

V. Normative Grounding and Self-creation

(a) Priority of Created Self

When an agent creates herself, what is the normative grounding relation between the self doing the creating and the self she creates? Just as one can ask whether intention or action is the ultimate locus of moral worth, one can ask whether S1, the creator self, or S2, the created self, is the ultimate locus of value. Theorists of self-endorsement and theorists of self-cultivation agree in taking S1 as the normative ground of S2. For the theorist of self-endorsement, S1 and S2 represent divisions within a person at a given time. S1 is one’s evaluative system, and S2 is whatever feature (or prospective feature) of oneself one is using that system to evaluate. If S2 is endorsed, this is *because* S1 has done the endorsing. For the theorist of self-cultivation, S1 and S2 are different time slices of the person. If the person has cultivated a feature in herself, then S2 should have that feature *because* S1 did something—made a commitment or a resolution—picking that feature out as the one to be acquired.

The normative priority of S1 comes through particularly clearly in Joseph Raz’s discussion of self-creation. Raz describes the earlier self as having done something akin to making a promise, and the later self as therefore having a new reason for action that is akin to one’s reason to keep a promise. Raz says that

“Our life comprises the pursuit of various goals, and that means that it is sensitive to our past. Having embraced certain goals and commitments we create new ways of succeeding and new ways of failing. In embracing goals and commitments, in coming to care about one thing or another, one progressively gives shape to one’s life, determines what would count as a successful life and what would be a failure. One creates values, generates, through one’s developing commitments and pursuits, reasons which transcend the reasons one had for undertaking one’s commitments and pursuits. In that way a person’s life is (in part) of his own making. It is a normative creation, a creation of new values and reasons. It is the way our past forms the reasons which apply to us at present. But it is not like the change of reasons which is occasioned by loss of strength through age, or the absence of money due to past extravagances. Rather it is like the change occasioned by promising: a creation, in that case, of a duty one did not have before. For, whatever reasons one had to make the promise, its making transforms one’s reasons, creating a

new reason not previously there. Similarly, the fact that one embraced goals and pursuits and has come to care about certain relationships and projects is a change not in the physical or mental circumstance in which one finds oneself, but in one's normative situation. It is the creation of one's life through the creation of reasons.”(p.387)

In Razian self-creation, S1 holds the reins, creating values and reasons for S2. Raz responds to the objection that such creation is arbitrary by leaning on the analogy with promising: “It may have been wrong to promise to give my son fireworks, for they are too dangerous. But having made the promise it may now be my duty to give him the fireworks.” This analogy plays right into Strawson’s hands: *why* did the father make that promise (i.e., create those values)? Had a yet earlier self promised to make these promises? Or did he promise for no reason, arbitrarily? Raz’s vulnerability to Strawson’s dilemma is, I propose, a product of his understanding of the normative grounding relation between S1 and S2. Promising exemplifies a form of agency in which the earlier self is normatively prior to the later one, and Raz takes this structure to be present in self-development: one’s earlier self makes it the case that certain norms govern one’s later self.

Let us review the dialectic thus far. We began by considering whether someone could create herself by endorsing or affirming some (prospective) value. We found that if a person stepped back far enough to avoid simply reaffirming what she already values, her endorsements no longer constituted value-generation. The self that is in a position to fairly, impartially evaluate its own values issues not values but, at best, judgments as to rational entailments amongst values. But perhaps a judgment as to what one’s current set of values entails could form an initial stage in the process of self-creation? From this question emerges the self-cultivation model, which describes the process of following through on such a judgment by making oneself into the person one judged one ought to be. But self-cultivation is vulnerable to Strawson’s dilemma, in response to which we found ourselves in a regress, stepping backwards in an attempt to identify ever earlier selves from which an act of self-creation might (non-arbitrarily) spring.

The problems seem to emerge from the fact that we press each model backwards, asking “who endorses the endorser?,” and “who cultivates the cultivator?” Why are we trying to go backwards in the first place? The answer, I believe, is the direction of grounding. Both models make S1 normatively prior to S2. The regressive line of questioning seeks S1’s normative advantage over S2, which is to say, the source of S1’s authority to endorse, or make commitments on behalf of, S2.

It is important to note that Strawson’s dilemma does not indicate a problem with the very idea of self-endorsement or self-cultivation. The theorist of self-endorsement can explain why the evaluating self has normative authority over the self it evaluates: the former simply *is* the locus of value. It is in the nature of one’s values that they are in a position to evaluatively endorse or reject our desires, motives, habits, etc.—but not our values. For those items are *members* of S1. Only in the case of value-endorsement does the demand to ‘step back’ from our values find application.

Likewise, self-cultivation is an intelligible way to make sense of efforts at self-change. Suppose S1 concludes, after careful reflection, that he ought to quit smoking. Holton can

explain why this resolution not to smoke binds S2 when the temptation to smoke strikes. S2 has reason to enact—without reconsidering the issue—the judgment that he made under rationally favorable conditions. We form resolutions (in part) because we anticipate the phenomenon Holton calls “judgment-shift,” in which occurrent desire sways our judgment as to what we should do. S1 was in a better position to adjudicate S2’s decision than S2 is. The problem is that in the case in which values are successfully created, S2 is always be in a superior valuational position to S1.

The theorist of self-cultivation is left grounding S1’s authority in the act of cultivation itself. But a vicious regress results from explaining the cultivating self’s rule-making authority by appeal to her status as cultivator. For she will, in turn, have to derive that license from a yet earlier cultivator, and we will be hurled backwards toward an ever receding source of ultimate authority. I propose that this regress reflects the fact that, in the case of self-creation, S1 cannot do the work that self-endorsement or self-cultivation would require of her. She does not have the normativity she is supposed to bestow.

I submit that the theorist of self-creation needs to get the creator self looking forwards rather than backwards: instead of imagining my future self as beholden to my past self, I suggest we imagine my past self as looking forward, trying to live up the person she hopes to become. The creator self doesn’t *make* a promise, she sees (to take up another facet of the concept of a promise) a *promise* of a better self. When we speak of some prospect as a promising one, we do not use the word ‘promise’ literally, since, among other reasons, what doesn’t exist yet cannot make promises. But this suggestive locution captures the reversal I propose. Promising presupposes a certain stability and predictability in one’s self and one’s circumstances. In a case where the values in question are in need of being created, it is only S2 and not S1 who would be in a position to take on such commitments. It is not S1’s place to embrace goals on behalf of, form reasons for, or create duties that will bind the created self. S1 is, by her own reckoning, not as she evaluatively should be. S1 is not the lender but the borrower of normative authority.

My proposal, then, is that we reverse the grounding relations between the two selves: S1’s decisions, feelings and actions in relation to the value in question derive their legitimacy from the fact that they represent a process by which she will arrive at S2. This move allows us to meet the demand contained in each horn of the dilemma without falling into a regress. The one horn—no random changes—demands normative continuity between the two selves, while the other—no derivable values—demands novelty in the created self. We can represent these demands thus:

The Continuity Requirement says that S1 and S2 stand in some normative grounding relationship to one another in respect of the value

The Novelty Requirement says that S2 must contain a value or values not grounded in the values of S1.

The continuity requirement and the novelty requirement only contradict one another if we fill out the former by assuming that S2’s values are grounded in those of S1. In the next

section, I want to flesh out what abandoning this assumption looks like. We will find that when we allow for the possibility that a process of value acquisition progresses toward its own source of normativity, Strawson's dichotomy gives way to a third option: that the creator self relate to the self she creates aspirationally.

(b) Self-Creation by Aspiration

Consider how people come to appreciate the value of classical music or religious observance or fashionable dress or fine cuisine or political debate. The process might begin because one is suddenly, experientially confronted with a value one's prior valuations did not lead one to anticipate. One might experience such a transformative moment at a performance, or in a church, or among new friends. Such experiences are 'transformative' only in the inchoative sense. They will not bear the weight we will have to put on them, if we try to use them to extrapolate a commitment to the endpoint. The transformed agent will look back, years later, and say of that initial event 'little did I know, back then, what was really valuable about classical music' (or religious observance, or fashion, etc.). She can, at that point, say "*this* was what I was after all along," but only because she has encountered "this" before knowing exactly what "this" was. The value she comes to endorse is one she knows as a result of working toward a target she could not, at the time, exactly envision.

All of us get some of our values from that early-childhood period in which our ethical development was managed by others. Perhaps the Strawsonian will want to say, with respect to these values, that we simply find ourselves with interests and mastery we played no role in generating. And then she will tell us that we can work up additional values by rationally developing the commitments implicit in them. Alternatively, we can once again be passively subject to environmental influence and acquire values that bear no rational relation to those already inculcated in us. But are these the only choices? For all of us have developed passions—for fine food, politics, music, or philosophy—long after leaving our parents' homes. And if we consider the course of such development, we find, I submit, that the Strawsonian picture simply doesn't ring true.

Given the expertise and work involved, it is implausible that anything but the earliest stages of such a transformation can be explained through fully external factors. For instance: the fact that someone found himself, for incidental reasons, in the exceptional gastronomic environment of Osaka, Japan might be the beginning of the story. Those experiences could ignite a spark of interest, but then something more would be needed to drive someone's systematic development of that initial spark into a full-fledged passion. The 'something more' in question is unlikely to be a value to which she was antecedently committed, from which a passionate interest in culinary excellence could be derived.

I don't want to deny that there might be cases in which an accidental and transformative initial experience or an anterior value commitment suffices to explain someone's value transition. But in many cases, we embark on these sorts of adventures without thinking that we know, in advance, exactly what we will get at the end. And that is not to say that we take them up on a whim, or for no reason, or by accident. There is an intermediate possibility that Strawson's dilemma has directed us away from recognizing, the

possibility that someone has an inkling of a value he does not fully grasp. He doesn't have a fully worked out sense of how this value fits into the rest of his values, because he doesn't have a fully worked out sense of what this value is. How could he, if the value corresponds to the intrinsic pleasures of the fine discriminations he is not yet capable of making? Nor is his pursuit conditioned on the coherence between the new value and the rest of what he currently values. Indeed, his ardent pursuit of it may take him away from much of what he currently cares about.

Most of the profoundly important activities, relationships, and forms of knowledge that human beings pursue are ones a person can fully appreciate, and integrate into her value-system, only once she is well-acquainted with them. And our question concerns the process of becoming acquainted. If you had to acquire values either by accident or by working through the entailments of your prior commitments, there would be no such process. You would either already be, in effect, at your value-destination, or have no way to get started. But those don't seem to be our only options.

The way in which people stand toward many of the values that they do not fully appreciate is that they partly appreciate them. And with respect to some of these partly-appreciated values, they also have the inclination to appreciate them more. They have a sense that their inchoate appreciation is *incomplete*, and act in order to attain a better valuation condition. The actions they perform are versions of the actions they will be in a position to perform once they have fully acquired the value: i.e., the one who wants to value music acts like the person who already loves music. But she also acts *unlike* the person who already loves music, since that person is not herself acting like anyone else. Such people are, in effect, imitating or trying to live up to someone. They don't pre-approve of the person that they're trying to be; rather, they hope that that person they aspire to be would (and will!) approve of them.

If you are trying to get better acquainted with some value, then you take your antecedent conception of that value to be inadequate. You act in order to grasp the value better, but your reason for wanting to grasp that value must be the very value you don't yet fully grasp. Life is full of moments in which one contemplates some obscure value from a great distance. We can't comprehend the value of child-raising for us, let alone the value of the life of the child we will raise, before starting a family. We go to college for the education college will itself teach us to appreciate; we leave our hometown with the aim of making some foreign place home; we date in order to love, and get married in order to love in a new way; we choose a career because of the as yet unfamiliar joys of expertly doing the work in question. In pursuing these values, our attitude is not merely a hope or wish that we will one day come to appreciate them. We *work* to appreciate them, and this work is rationalized and guided by the value we are coming to know. In these cases, the full justification of what we are doing can only come at the end of the story. It is the end that provides the normative standards for assessing what comes before it.

But why think of the aspirant as being under the guidance of her future self, as opposed to being under the guidance of the very value to which her future self is (more) responsive? It is not wrong to say that the aspirant is guided by the value she shares with her future

self, but this guidance must nonetheless be filtered through that self. For consider how many values we apprehend the existence of, without trying to change ourselves to acquire. We can only dedicate ourselves to so many self-changing pursuits. Thus we aspire with respect to a small subset of potential values, ones with reference to which we can anticipate that a successful grasp is in the cards for us. These are the cases in which we can (dimly) see our way to the endpoint. And this is just to say that being able to envision the person we will become when we have the value is necessary for aspiration. (Recall Dickens' Sidney Carton.) Moreover, with respect to a given value, there are often myriad ways in which one might come to value it. Culinary appreciation might take the form of learning to cook, or frequenting restaurants, or becoming a chef, or growing one's own vegetables. When I aspire, I work my way toward some particular mode of appreciating the value, namely the mode I eventually end up with. This is not to say that I can foresee in all its details, or that I won't make mistakes, but that my efforts are directed at something more concrete than just the goal of food appreciation. It is not the goal of food appreciation considered in the abstract, but the particular form of it that I am working to develop, that sets the standard for what makes a given turn in my development count as mistaken or correct. Thus it is, after all, correct to say that I am working to make myself into the particular gourmand that I become. I turn now to address the paradoxicality of that formulation.

VI. Teleology and Agency

(a) Normative Grounding and Backwards Causation

The aspirational account of self-creation offers us a recognizable story of what it is like to actively acquire values, and avoids the paradoxes and problems associated with trying to force our understanding of that transition into the mold of self-endorsement or self-cultivation. It may nonetheless arouse suspicions due to its teleological character. We are not, generally, inclined to accept forms of explanation in which that which is later grounds what is earlier. As a first pass at allaying this worry, we might consider the fact that intentional action is generally taken to have a teleological character. When I act, there is (usually) some state of affairs, the subsequent occurrence of which constitutes the action as a success. My action at t_1 can be assessed in the light of whether or not, at t_2 , this state of affairs obtains. Insofar as we are prepared to allow that actions can succeed or fail with reference to the obtaining or non-obtaining of the end at which they are directed, we take actions to be normatively grounded in what they bring about⁹⁴.

But this vindication of teleology will not suffice. For our interlocutor may insist that the end normatively grounding the action itself be normatively grounded in something temporally earlier. She holds that an action of ϕ -ing can be assessed in the light of end E only because E, in turn, figured in the content of the intention with which that action was performed: when the agent ϕ -ed, he was trying to bring about E. Furthermore, she claims, that intention serves as a ground for assessment of the end because it can, in turn,

⁹⁴ Though perhaps not everyone is prepared to allow this. See my sketch of naïve Kantianism in IV of this chapter.

be assessed—as rational or irrational—in the light of the agent’s beliefs and desires. Thus, on her account, my ϕ -ing can (teleologically) satisfy E only because E itself satisfies my antecedent intention, which in turn satisfies the my antecedent beliefs and desires. On this picture, the *ultimate* normative ground in the light of which ϕ -ing can be assessed is the agent’s set of antecedent conative and cognitive commitments. Someone might grant that intentional action is teleological in that it can be assessed as succeeding or failing in the light of what results from it, but insist that this teleological feature is predicated on a non-teleological conception of the normative relationship between the end and the agent’s antecedent intention and motivation.

By contrast, the normative standard by which we assess aspiration does not show up in its complete form until the aspirational activity is completed. It cannot be read off the agent’s antecedent beliefs and desires unless those beliefs and desires are, in turn, interpreted proleptically, i.e., in relation to what comes after them. The end for the sake of which the agent acts when she aspires is not itself to be assessed in the light of what she currently takes herself to be doing, or the desires with which she entered into the pursuit. Rather, the order is reversed: both the intentions with which she acts and the motivations from which those intentions spring are to be assessed in terms of their conduciveness to her aspirational end. If the desires she has are not of the kind that characterize the person she’s trying to become, she wants different desires.

The opponent of teleology may suspect that I am invoking something like backwards causation or time travel: the future cannot cause the past, so how can something that is *in* the future have an effect on what I am doing *now*? Unless the ground of the agent’s self-transformation is there at the outset, how is she to move herself to the desired result? Teleology, it may be claimed, must be underwritten by—i.e., grounded in—something that is temporally prior. It is at this point that we must return to a distinction, noted in passing in the previous chapter, between intra-normative and extra-normative grounding relations.

As we observed, theorists of grounding are apt to distinguish between the normative grounding claims characteristic of ethics and the natural grounding claims characteristic of science. Since the latter describe relations of efficient causation, they can never have a teleological structure. Efficient causes must be prior to their effects. But of course ethicists can and do ask questions about the natural grounding of action. We should not, therefore, confuse the claim that a teleological ground (such as S2) must be *naturally* grounded in something temporally prior with the claim that it must be *normatively* grounded in something temporally prior.

When an agent acts on a proleptic reason, she takes steps toward bringing about the normative ground of her present condition. She works to make herself grasp the reason (R2) that is, normatively speaking, the full or complete or nondefective variant of the proleptic reason (R1) causally driving that work. R1 is the causal ground of R2, though R2 is the normative ground of R1. R1 brings about R2 even though R2 offers us the standard by which to assess the rightness or wrongness of R1. As an example of another case in which causal priority fails to track normative priority, consider the case of

predictions. A prediction made at t1 about an event at t2 is to be assessed only at t2⁹⁵. My prediction is correct if and only if the event I predicted occurs. But of course my prediction is temporally prior to the event I predict. Indeed, my prediction might even contribute to the causal genesis of the event I predict, as in the case where, e.g., someone's prediction that she will perform badly causes her to lose confidence in herself and, ultimately, to satisfy the prediction. The prediction must be assessed as correct or incorrect on the basis of her subsequent performance.

But this is exactly the sort of example that will worry the theorist of action. For he will point out that unless my prediction was responsive to norms I could apprehend at the time of making it, my success cannot be ascribed to me as something that calls for, e.g., epistemic praise. If I just *guessed*, then the correctness of my prediction does not represent a genuine achievement of mine. Likewise, in the practical domain, we can say that in order for some behavior to qualify as a phenomenon or manifestation of agency, it must be in some way the agent's achievement. Christine Korsgaard has expressed this in terms of what we might call a "guidance condition" on rational agency: "A person acts rationally...only when her action is the expression of her own mental activity..." (1996, p.33); "A rational agent would be guided by reason in the choice of her actions." The action theorist may argue that in order for some behavior to be assessable as succeeding *in a distinctively practically rational way*, the norm with reference to which she is supposed to have succeeded must be mentally present to her at the time of action. For only in that way could the norm have guided her action. In order to understand aspiration as the aspirant's own work, we need to be able to understand it as coming *from her*. Or so she maintains.

We can, then, set aside more general worries about teleology and focus our attention on the worry specific to the action theorist: can we represent the aspirant's progress as something she does, given that she lacks a clear or determinate conception of her target until she arrived at it? Does the aspirant satisfy Korsgaard's guidance condition?

(b) Two Conceptions of Agency

If Davidson's causal theory of action, read a certain way, is correct, she will not. Suppose that some behavior counts as agency in virtue of the fact that it was caused by a belief and desire that rationalize it. On this picture, rational guidance is a causal matter. So, for instance, my trip to the store is caused by the desire to get food and the belief that I can get food at the store. My trip succeeds if I do indeed end up getting food at the store, and it can be assessed in the light of this norm precisely because this norm played a causal role in the genesis of the trip. On at least one way of understanding the desire that functions in such a causal explanation, it must constitute a grasp of just what is desirable about the outcome. Some event counts as agency to the extent that its success conditions are inscribed in the attitudes causing it. If practically rational guidance required an agent to know exactly what she wanted out of the outcome, aspiration could not qualify as rationally guided. The aspirant fails to grasp the full normative grounding of her project

⁹⁵ I am referring here only to assessment in terms of correctness, rather than in terms of how rationally the agent treated the available evidence.

until it is completed. And if the reason is fully available to the agent only at the conclusion of the aspiration, it cannot (efficiently) cause the behavior that constitutes her aspiration.

A number of philosophers have recently developed a line of thought from Anscombe into a rejection of the causal theory of action. The alternative can be expressed as a response to the well-known problem of ‘deviant causal chains.’ The worry is that the Davidsonian has no way of ensuring that it is really the reason, as opposed to something correlated with the presence of the reason, that is doing the relevant causal work⁹⁶. So, for instance,

“A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never chose to loosen his hold, nor did he do it intentionally.” (Davidson, “Freedom to Act” p.79 fr. *Essays on Actions and Events*)

Davidson says that he himself “despair[s] of spelling out...the way in which attitudes must cause actions if they are to rationalize the action.” (p.79) Some, perhaps ultimately including Davidson himself⁹⁷, have taken the problem as fatal to the causal theory. If we cannot eliminate wrong causal chains, then we may be inclined to give up on the project of basing our answer to the question of whether some event is an action on the way in which that event was caused.

Some of Davidson’s critics⁹⁸ have sought insight in Elizabeth Anscombe’s *Intention*. According to the account they find in Anscombe, some behavior counts as an action—and therefore as assessable in the light of some end—not in virtue of the beliefs and desires that constitute its cause, but rather in virtue of the fact that, throughout the course of the action, the agent has a distinctive practical knowledge of what she is doing. So, some behavior counts as, e.g., a hand-raising done for the sake of voting, because the agent does it in the practical knowledge that that is precisely what she is doing. This kind of knowledge is distinctive in being non-observational and non-inferential. When *I* try to answer the question as to what she is doing, I must observe her behavior and perceive that her hand is going up. In order to know that she is raising her hand in order to vote on the proposal, I must draw an inference from that (and other) observations. When *she* tries

⁹⁶ This famous worry can be articulated negatively, as to how one rules out deviant causal chains ([Frankfurt (1978). “The Problem of Action.” p.43 in Mele reprint] [Davidson 1960, 153-4) or, positively, as to whether the semantic element is doing the causal work (Dretske 1989). For a recent defense of Davidson on this point against a rival neo-Anscombean picture, see Paul 2011. For an argument that the deviancy point needn’t be answered, see Torbjörn Tännsjö’s “On deviant causal chains – no need for a general criterion.”

⁹⁷ Davidson 2004, “Problems of Rationality”: 106 (“Several clever philosophers have tried to show how to eliminate the deviant causal chains, but I remain convinced that the concepts of event, cause and intention are inadequate to account for intentional action.”)

⁹⁸ Here I rely on Sarah Paul’s drawing together of work by Michael Thompson, Sebastian Rödl, Candace Vogler, Richard Moran, Martin Stone, Matthew Boyle, Douglas Lavin, Kieran Setiya and J. David Velleman. She helpfully abstracts the differences between their views to isolate an Anscombean emphasis on the formal, as opposed to efficient, causal structure of action. (cf. Paul fn. 12)

to answer the question as to what she is doing, she needn't observe herself or draw any inference. Thus observational or inferential knowledge of what someone is doing contrasts with the immediate, first-personal knowledge characteristic of the agent acting. Does the Anscombean account of agency offer better resources than the Davidsonian theory to the theorist who is accounting for aspirational agency?

It does not. If the agent's reason is to be the object of her practical knowledge (or even her practical belief) she must fully grasp it in advance of the action's coming to an end. And this is just what the aspirant cannot do. She does not know, or even take herself to know, what she is doing. She is dissatisfied by her own answer to the "why?" question. Whether we make the agent's reason for action the efficient cause of the action or the object of the agent's practical knowledge, we seem to presuppose on the part of the agent an access to the normative standards governing her action. But the aspirant lacks this access, being unable to fully articulate, either before or during her action, exactly what she is doing and why. This is not to say that she has *no* idea what she is doing, but rather that her conception of what she is doing is derivative of a superior conception she will have after she has arrived at her destination. The grasp of what she is doing that (fully) explains what she is doing is not to be found until her agency is complete. Thus we cannot say, of aspirational agency, that the normative assessability of its results derives from the fact that some conception of those results was present to the agent during the course of the agency. In the case of aspiration, the representation in question will have to be assessed in the light of another representation, namely the representation that *is* the result. If we want to ground the *normativity* of aspiration in antecedent representation we will run in a circle.

One way of accommodating either the Davidsonian or the Anscombian approach would be to break up the aspiration into its many component actions. We might present each of those actions as caused by, or done with the knowledge of, some reason that rationalized it independently of the value that stands at the end of the process as a whole. On such a piecemeal picture, the many actions add up to a change of value at the endpoint, but that change of value is not a target of her agency. Indeed, the proximate face of someone's proleptic reason will provide us materials for this sort of reduction: we can say that I listened to the opera because I wanted a good grade; I wanted to please my parents; I promised myself a chocolate bar when I got to the end. But we have seen, in chapter two, that such a picture of aspiration will not do: the theorist of aspiration must supply a rationalization of the process of aspiration as a whole. Treating each action in isolation from the ultimate aspirational target produces a distortion of what the agent is doing; on the right way of looking at these moments of agency, they are not self-standing.

(c) Aspirational Agency vs. Self-Standing Action

It is, perhaps, no flaw in the Anscombian or Davidsonian account that they do not accommodate aspirational agency. For though some actions (e.g. building a cathedral) may take a long time, it does not seem natural to characterize the whole process of becoming a music-lover, or a parent, as a single action. It is easier to apply that label to the smaller projects—such as taking a music class, or outfitting the nursery. But if I am

right, those actions will, insofar as they form part of aspiration, be of a distinctive kind: they will be actions that cannot be understood except with reference to a larger stretch of agency of which they are a part. Anscombe and Davidson were offering accounts of what we can now characterize as “self-standing actions.” The aspirant’s agency does not have this quality: her individual actions are not self-standing, and the aspirational whole of which they are a part is not an action. In addition to the self-standing agency of the clear-eyed agent described by Anscombe and Davidson, there is a form of agency appropriate to the one whose agency is a form of learning.

In non-proleptic (standard) practical reasoning, the agent looks to the world for a match for her antecedently fixed goal. For instance, she goes to the store in order to get milk for her cereal. If the store is closed, she will be disappointed—and she knows this in advance. Which is to say, she knows in advance what will constitute the success of her action, and what failure. She can say exactly why she is doing what she is doing. The proleptic reasoner, by contrast, does not demand that the end result match a schema she currently claims in her possession. For she wouldn’t know what to ask for: she hopes, eventually, to get more out of what she is doing than she can yet conceive of. Proleptic rationality illustrates the possibility that we might engage in practical rationality in order to improve the very conception of the good that drives our reasoning.

Where does this leave us with respect to self-creation? The question of whether self-creation is possible turns out to be the same question as the question of whether there are proleptic reasons. Self-creation, as I have been describing it, relies on the possibility of being guided by a grasp of a reason that is inchoate in the way described in chapter two. We can now describe this inchoateness as a matter of normative dependence: her grasp of the reason is normatively grounded in the grasp she will have once she becomes the person she’s trying to be. Does aspiration, thus described, satisfy Korsgaard’s guidance condition?

Let me spell out that condition as follows: in order for some stretch of agency done at time t to be done for the sake of norm N , it must be true that the agent grasps N at t . This statement of the condition leaves open the explanatory relation in which these two facts—the fact that the behavior at t is for the sake of N , and the fact that the agent grasps N at t —are meant to stand. Suppose the idea is that the second does all of the explanatory work, while the first plays the role of what is to be explained. In that case, only what I have called “self-standing” action will satisfy the guidance condition. For only in those cases will the condition at the endpoint be explained by, and not be explanatory of, the antecedent grasp the agent has of that endpoint. In cases of aspiration the direction of explanation runs both ways. In order to explain the fact that the agent’s arrival at the endpoint is to be ascribed to her as her own work, we need to cite the grasp she had at the time of action. But when we do so, we are citing a merely proleptic grasp, namely one that must itself be explained as an attenuated version of the grasp that constitutes her condition at the endpoint.

All instances of agency are teleological, in that they are done for the sake of some subsequent end. The agent’s behavior is “going somewhere,” and that is because it is

guided, from within, by the agent's sense of where she is going. What makes the aspirational cases special—puzzling—is that they are cases in which the agent's sense of where she is going is also, in the relevant sense, “going somewhere.” The teleological structure of aspirational agency is, in this way, more robust than the teleology of self-standing actions.

Chapter 6: Implications of the Aspirational account of Self-Creation

In this chapter, I propose to display the fertility of the aspirational account of self-creation by deriving various applications. I first apply it in a historical way, in the service of analyzing Aristotle's account of virtue. In reading Aristotle's claim that one becomes virtuous by doing virtuous actions, interpreters face a version of Strawson's dilemma. On one interpretation, Aristotle is saying that the person coming to be virtuous does actions that have merely the external aspect of virtuous actions. On another, the actions of the one being habituated are already done "for their own sakes," and from the correct (moral) motivation. The first interpretation fails to recognize that the virtue we end up with must be continuous with the virtue we develop in ourselves, while the second ignores the novelty requirement⁹⁹. I show how the aspirational account of self-creation fits Aristotle's text better than either of the two standing interpretative options. Historians of philosophy have no more escaped the bind of Strawson's dilemma than their counterparts outside of history; in exposing a way out of the dilemma, we become not only better able to understand self-creation, but also better able to read Aristotle.

In the second part of this chapter, I turn to discuss the implications of the aspiration theory for questions of moral responsibility for one's valuational condition. First, I clarify the aspirational account by distinguishing aspiration both from the "dialectical activity" described by Talbot Brewer, and from the phenomenon of ambition. I argue that the aspirational account of responsibility is *asymmetrical*, accounting for a person's responsibility for acquiring a good character in a different way from her responsibility for acquiring a bad one. I end by using a case study to compare the explanatory power of the aspirational account of responsibility with that of the self-endorsement view, the self-cultivation view, and Susan Wolf's "Sanity" view.

I. Aristotle on Habituation

As an illustration of the power of the aspirational approach, I want to exploit it in interpreting an author who is concerned to give an account of rational development, and who approaches the problem with a different set of assumptions from our own. What we will see is that the aspiration theory frees us from the stranglehold of two bad interpretative options.

Aristotle begins chapter four of the second book of the *Nicomachean Ethics* (NE) with as explicit an acknowledgment of paradox as we find anywhere in the corpus:

The question might be asked, what we mean by saying that we must become just by doing just acts, and temperate by doing moderate acts; for if men do just and moderate acts, they are already just and moderate, exactly as, if they do what is grammatical or musical, they are proficient in grammar and music. (1105a17-21)¹⁰⁰

⁹⁹ I refer here to the continuity and novelty requirements from ch. 5, part V, a.

¹⁰⁰ Translations of Aristotle's *Nicomachean Ethics* are from Broadie 2002, Translations of all other works of Aristotle are from Barnes 1984.

The paradox arises because Aristotle holds two apparently conflicting theses:

The Virtue Acquisition Principle: people become just by doing just and moderate actions.

The Virtue Activation Principle: just and moderate actions activate, and therefore presuppose the existence of, a person's justice and moderation.

The Virtue Activation Principle is a truism about the relation between a disposition (hexis) and the activity in which the disposition is manifested. Justice and moderation are dispositions of character, which is to say that they pick out the states from which a person performs the corresponding actions. Justice is the power to act well, so acting justly presupposes justice in the same way that a given act of seeing presupposes the power¹⁰¹ to see.

But why does Aristotle hold the virtue acquisition principle? Aristotle is well known for asserting that virtue is a product of habituation. It may, then, come as some surprise that he finds himself in the circle described above. Why doesn't he simply say that you acquire virtue from your parents? It will be helpful to look at the previous chapters of the NE in order to get the background necessary for understanding both the virtue acquisition principle and the true nature of what goes by the label "habituation."

Book I ends with a distinction between the affective part of the soul, which Aristotle takes to be concerned with pleasure, emotion, desire, and feelings (pathē) generally, and the intellectual part of the soul, which is responsible for both practical and theoretical thinking. Aristotle makes a corresponding distinction within virtue, identifying forms of excellence appropriate to each part of the soul. The virtues of the affective part are the familiar list of ethical virtues. The list includes the two mentioned in the quote above, justice and moderation, as well as courage, generosity, gentleness, great-souledness, wittiness etc. The excellence of the intellectual part is wisdom, which Aristotle distinguishes into a practical form (phronēsis) and a theoretical form (sophia). In book II he points out that we possess neither the virtues themselves, nor the capacities for them, by nature. How, then, are they acquired? He sets aside the intellectual part of the soul, which is perfected by teaching, and turns his attention to the process by which one acquires ethical virtue. Aristotle's first observation about this process draws an etymological connection between the ethical (ēthikē) quality of states of character such as courage, justice and moderation, and the word, ethos, describing the process by which they are acquired: "...excellence of character [ēthikē] results from [habituation] *ethos*—which is in fact the source of the name it has acquired, the word for 'character trait' [ēthos] being a slight variation of *ethos*." (1103a17-18)

The idea that the ethical virtues are acquired by the process in question strikes Aristotle as a truism illustrated by the etymological connection between the words themselves—ethos is whatever process gives rise to what is ethical. This fact can easily be obscured by the standard translation of the various words Aristotle uses to refer to this process

¹⁰¹ Dispositions (hexeis) are distinct from powers (dunameis) in that the former are acquired, while the latter are present by nature. It is only because he takes virtue not to be a power that Aristotle is faced with this paradox.

(*ethos*, *ethismos*, *ethizesthai*), which is “habituation.” The translation is, in and of itself, not a bad match for at least one of the Greek words Aristotle uses to refer to the process, namely *ethizesthai*, but it can, for two reasons, obscure Aristotle’s distinctive conception of the process in question¹⁰². First, unlike Aristotle’s words, it bears no internal conceptual connection to the word we use to translate what it generates, namely “(ethical) virtue.” Thus, the claim that an ethical action come from the ethical disposition (i.e. just actions come from justice) will strike us as much more straightforward than the claim that the ethical disposition comes from habituation. But for Aristotle these are close to parallel.

Second, and more importantly, the word “habituation” refers to a process in which the parent or teacher *acts upon* the child or pupil. To habituate someone is to do something *to* them, and to be habituated is to undergo some experience of training at the hands of another. There is no active verb that describes what the subject of habituation or training does, since the active role is reserved for the teacher or parent or mentor or coach. This feature of the English is not lexically inaccurate, since the Greek word *ethizesthai* has the same structure: it describes a process of which the developing youth would be the patient. Nonetheless, Aristotle is very clear that the process as he conceives it does not mirror this linguistic fact. For he says in many places that it consists of actions (*praxeis*) performed by the person acquiring virtue (or, as I will call her, the “trainee”):

...we acquire the excellences through having first engaged in the activities, as is also the case with the various sorts of expert knowledge—for the way we learn the things we should do, knowing how to do them, is by doing them. For example people become builders by building, and cithara-players by playing the cithara; so too, then, we become just by doing just things, moderate by doing moderate things, and courageous by doing courageous things. (1103a31-b2)

I am not suggesting that we avoid translating *ethos*, *ethismos*, *ethizesthai* as “habituation.” My point is only that we must attend to the fact that Aristotle, working with the linguistic resources of Greek, plays up some of the linguistic features of the word—such as its etymological connection with character—and downplays others—such as its allocation of the passive role to the trainee. Keeping track of what Aristotle is doing with the word “*ethizesthai*” will help us understand how his distinctive conception of habituation gives rise to the paradox. Keeping the standard translation, we must emphasize the fact that habituation has an internal connection with the disposition it produces, and that habituation consists in work done by the habituee. Nor are these two facts unrelated. Aristotle seems to be combining them when he insists that virtue comes about not through just any form of activity, but specifically through a form of activity that is *akin* to that which it produces:

Again, it is from the same things and through the same things that every excellence is both produced and destroyed, and similarly every expertise; for it is from playing the cithara that both the good and the bad cithara-players come about. So too both with builders and the rest: good building will result in good builders, bad building in bad ones. (1103b6-12)

¹⁰² There is also the problem that it suggests that what is arrived at are mere ‘habits,’ that is, rote behavior ingrained through sheer repetition. It has been well observed in the secondary literature [REF Burnyeat, Kosman] that this is a mischaracterization of the activity of the virtuous agent: when he acts from his character, he is not merely doing what comes habitually.

Aristotle is not blind to the fact that the process he is describing has a dependent character, relying on external support. He makes reference to the need for a teacher whose job is, presumably, to guide, assist and correct the learner: “If it were not like this, there would be no need at all of anyone to teach them, and instead everyone would just become a good builder or a bad one.” (1103b12-13)

Finally, he concludes the chapter by turning his attention specifically to the ethical (as opposed to craft) cases, and pointing out that the process in question involves reshaping one’s passions:

This, then, is how it is with the excellences too; for it is through acting as we do in our dealings with human beings that some of us become just and others unjust, and through acting as we do in frightening situations, and through becoming habituated to fearing or being confident, that some of us become courageous and some of us cowardly. A similar thing holds, too, with situations relating to the appetites, and with those relating to temper: some people become moderate and mild-tempered, others self-indulgent and irascible, the one group as a result of behaving one way in such circumstances, the other as a result of behaving another way. We may sum up by saying just that dispositions come about from activities of a similar sort. This is why it is necessary to ensure that the activities be of a certain quality; for the varieties of these are reflected in the dispositions. So it does not make a small difference whether people are habituated (*ethizesthai*) to behave in one way or in another way from youth on, but a very great one; or rather, it makes all the difference in the world. (1103b13-25)

Taken out of context, as it often is, this last sentence could easily be heard as a reference to the importance of parents in shaping or even creating a child’s character. But that is not Aristotle’s point. His point is that we become the people we make ourselves. Without suggesting any retranslation, I note that we can accurately gloss Aristotle’s point in that sentence using the word “practice.” That word has a helpful semantic profile: it is used not only for the preparatory work (“practice makes perfect”) but also for the thing itself (a medical practice). Aristotle is saying that it makes a big difference whether people *practice* behaving in one way or another from youth on. Aristotle does not want us to understand the process as one of *being* changed and transformed. Consider the criticism with which he concludes his discussion of the virtue-acquisition paradox in II.4.

But most people do not do these, but take refuge in theory and think they are being philosophers and will become good in this way, behaving somewhat like patients who listen attentively to their doctors, but do none of the things they are ordered to do. As the latter will not be made well in body by such a course of treatment, the former will not be made well in soul by such a course of philosophy. (IV.2, 1105a17-b18)¹⁰³

Burnyeat (p.70) is surely right to hear in this passage a rejection of Socrates’ intellectualized conception of virtue as knowledge. But what are Aristotle’s grounds for objecting to that doctrine here? Aristotle is cautioning his reader against the overly passive conception of virtue-acquisition that he feels attends an intellectualization of virtue. He seems to be grounding his objection to intellectualism, here, on the fact that the intellectualist cannot tell the right kind of story about virtue acquisition. Those who

¹⁰³ Translations of Aristotle throughout are from Broadie and Rowe.

take refuge in philosophy behave like patients who *listen* but do not *do*.¹⁰⁴ Even sick people, Aristotle points out, need to take their well-being into their own hands at some point. They cannot expect just to be made healthy by the work of the doctor. The same holds for those of us sick in soul: we should not think that our own justice is anyone else's doing.

The paradox Aristotle articulates in II.4 comes, then, from his understanding of the process of virtue-acquisition as one of performing the actions that are akin to the disposition one has. For him, the claim that justice must arise from just actions is as truistic as the claim that just actions must come from justice. Hence he sees that he is caught in a circle, to which he responds in the following way:

(a) Or is this not true even of the arts? It is possible to do something grammatical either by chance or under the guidance of another. A man will be proficient in grammar, then, only when he has both done something grammatical and done it grammatically; and this means doing it in accordance with the grammatical knowledge in himself.

(b) Again, the case of the arts and that of the excellences are not similar; for the products of the arts have their goodness in themselves, so that it is enough that they should have a certain character, but if the acts that are in accordance with the excellences have themselves a certain character it does not follow that they are done justly or temperately. The agent also must be in a certain condition when he does them; in the first place he must have knowledge, secondly he must choose the acts, and choose them for their own sakes, and thirdly his action must proceed from a firm and unchangeable character. These are not reckoned in as conditions of the possession of the arts, except the bare knowledge; but as a condition of the possession of the excellences, knowledge has little or no weight, while the other conditions count not for a little but for everything, i.e. the very conditions which result from often doing just and temperate acts.

(c) Actions, then, are called just and temperate when they are such as the just or the temperate man would do; but it is not the man who does these that is just and temperate, but the man who also does them *as* just and temperate men do them. It is well said, then, that it is by doing just acts that the just man is produced, and by doing temperate acts the temperate man; without doing these no one would have even a prospect of becoming good. (1105a21-b12)

There is some controversy over whether Aristotle is here presenting two or one solutions to the problem¹⁰⁵; and over the question of whether Aristotle is assimilating or contrasting craft learning and ethical learning. I propose that he offers a single solution in two stages, and that the first stage, (a), assimilates craft and ethics, while the second, (b), distinguishes them.

Aristotle opens, in (a), by drawing a distinction between doing an X-ish action, and doing it X-ishly. This is the basic structure for resolving the paradox: the acquisition of virtue X or craft X presupposes doing X-ish actions; when one has X, one not only does X-ish actions, but also does them X-ishly. Thus far Aristotle gives a single, unified solution to the paradox. However, Aristotle acknowledges that this solution is merely schematic, and must be filled out somewhat differently depending on the nature of X. The second

¹⁰⁴ Broadie, too, notes Aristotle's emphasis here on action as opposed to speech (p.301, comment to 1105b13-14) as a method of acquiring virtue.

¹⁰⁵ Broadie (1991), 119n.7 takes there to be 1 answer, whereas Irwin (1999), p.195 and Jimenez (2106) p.12-13 read the passage as providing two independent solutions.

stage, (b), presents us with the filling out, and offers details specific to the ethical case. In the final paragraph, (c), he returns to the abstract formulation of (a), this time using virtue instead of craft as his example.

The crucial passage is (b). What, exactly, does it mean to do a virtuous action but not do it virtuously? What is the “certain character”¹⁰⁶ (1105a29) which Aristotle describes the trainee’s just action as having, despite the fact that he doesn’t act justly? In the case of the crafts, doing the craft “X-ishly” means doing it with knowledge. In the case of virtue, “X-ishly” means fulfilling a set of three conditions: doing the virtuous action with knowledge, choosing it for its own sake, and having a fixed character¹⁰⁷. The first condition is a point of overlap with crafts, but Aristotle says that it is less important than the other two. The third, it might seem, simply restates the fact that the trainee is a trainee, and therefore must perform the action without (firmly) possessing the relevant character. This doesn’t appear very informative as to the question of how he nonetheless performs the relevant action. For these reasons, interpreters have tended to focus on the second condition, emphasizing the trainee’s failure to do the action for its own sake. The assumption is that the trainee is unmoved by the virtuousness of her action, and does it from an ulterior motive of reward, praise, fear, or the like.

Marta Jimenez (2016) summarizes the account of habituation that results from this line of interpretation:

It is common to see the learners’ actions characterized as being virtuous only ‘in an equivocal sense’ or *homonymōs* (Stewart), ‘only in their external aspect like those that virtue produces’ (Ross), not virtuous ‘in the same full sense as those which we do when our *hexis* is fully formed’ (Joachim), ‘not strictly virtuous’ or actions that ‘while not strictly virtuous or accomplished, have an external similarity to the performances which manifest a virtue’ (Hardie), ‘minimally virtuous actions’ and ‘acts that are less than fully V[irtuous]’ (Williams), ‘virtuous actions in a minimal sense’ (Vasiliou), and so on. In all these cases, scholars see the actions of the learners as lacking in relation to those of virtuous people because they are not done for virtuous motives—i.e. they are not done for the sake of the noble, but for some other motive. (p.6)

On this interpretation, the virtuous action can be picked out independently of the agent’s motive for engaging in it. For the trainee, the just action is that of returning what he

¹⁰⁶ *auta pōs echēi*

¹⁰⁷ There is another difference between craft and virtue mentioned by Aristotle in the first sentence of (b), namely a difference as to the very significance of the distinction Aristotle is drawing. In craft, the question of *whether* someone has merely done an X-ish action or also done it X-ishly is less important to us than in the case of virtue. This is because in craft our primary concern is neither the character nor the action but the separable product. We demand that the chair the craftsman makes for us be stable and comfortable, but, providing those demands are met, we are not inclined to also require that it be made by the master as opposed to the apprentice. This difference gives rise to a corresponding difference in the constraints on what will count as having done the action merely X-ishly. In the case of craft, even an accidentally perfect craft object can serve the relevant function. In the case of virtue, as I will argue below (see my discussion of V.8, fn. 109 corresponding text) the bar is raised. The action is not even going to count as just or moderate or courageous (let alone done justly or moderately or courageously) if it was done by accident. Thus, of Aristotle’s two initial examples of how one might do a grammatical action without the knowledge of grammar—namely, by accident and with the help of another—only the latter will be transferable to the case of virtue.

borrowed, the courageous one that of rushing eagerly into battle; temperance is refusing opportunities for sexual or gustatory pleasure¹⁰⁸. On this picture, his motivation for the action is irrelevant, so long as his action is in its “external aspect” like that of the virtuous person. I will, therefore, call this the “external aspect interpretation.”

The (fully) virtuous man would, then, be the one who does actions with the relevant “external aspect”—he returns what he borrowed, rushes into battle, and refuses pleasures—but only when they are in fact right and good thing to do. Unlike the trainee, he uses normative concepts to pick out a just or courageous or moderate action. He performs such an action not because it fits the description “money-returning” or “charging forwards” or “self-restraint” but because he sees that in the case in question he would be *rightly* returning money, charging forward, or restraining himself.

Jimenez is right to reject this picture of habituation on the grounds that it fails to allow for continuity between the agency of the trainee and that of the virtuous person. (p.4) It does not seem true that by doing many actions that fit the relevant non-normative description, one becomes just at the higher level. Repeatedly paying one’s debts, or rushing into battle, or forswearing pleasures, because, e.g., one is threatened if one does otherwise, does not lift one up into a disposition to do these things when they are the right thing to do. For the (fully) virtuous man will also *not* repay a debt, *not* rush into battle and *not* forswear pleasures, when *those* actions are the right thing to do. If the trainee’s *ethismos* (habituation) is to be a moral education, he must be learning precisely to answer the question, “when is it right to charge forward?” It seems he can hardly learn this by repeated blind charges.

There is also some textual evidence that speaks directly against this interpretation. When Aristotle says in V.8 that someone who pays a debt from fear of punishment neither acts justly nor does what is just, he seems to be making use of the distinction between two ‘levels’ of justice he draws in II.4.¹⁰⁹ He explicitly denies that someone whose action satisfied any non-normative description of justice—paying a debt—fulfills even the lower level of justice at which the trainee resides. We can, therefore, adduce this passage in support of Jimenez’ contention that in order for your action to be (in any sense) just, you must do it for the right reasons. When someone justly returns what he owes, he does so not in order to avoid a more painful outcome (much less from evil ends) but from some the recognition that it is the right thing to do. The ethical awareness that must

¹⁰⁸ Among the other problems with the proposed view lies the fact that these “external aspect” descriptions of courage and temperance avoid only one of the two vices between which the virtue is situated. The description I have given of courage could equally well describe recklessness, and the description I have given of temperance could describe the vice of insensateness. (Justice is special on this count, since it does not seem to be located between twin vices, but the problem is clearly present for all the other virtues.) Aristotle’s doctrine of virtue as a mean ensures that no nonnormative action description—which is to say, no action description that fails to invoke the right amount, at the right time, in the right way etc.—can pick out virtue.

¹⁰⁹ “...for a man might return a deposit unwillingly and from fear, and then he must not be said either to do what is just or to act justly, except in an incidental way.” (NE V.8, 1135b4-6) The final qualification does not save the interpretation being discussed. For Aristotle ought (if the interpretation were correct) to say that such a person unqualifiedly does what is just, though he does not act justly.

accompany (any) just action entails that no non-normative description can pick such an action out.

Let us, therefore, turn to consider Jimenez' interpretation. On her view, the virtuous action as done by the trainee is done from a virtuous motive of recognizing its ethical goodness. His actions can be called "just" and "brave" and "temperate" insofar as he returns money, rushes into battle or forswears pleasure when that is the right thing to do. On her view, the trainee virtue is a matter of "doing virtuous actions for their own sake and for their nobility." (p.28) I think this interpretation also fails. Where the first interpretation created too large a gulf between the trainee and the fully virtuous person, this interpretation both creates too small a gulf and raises the bar for trainee-hood problematically high.

On Jimenez' interpretation, we must wonder how one acquires the moral knowledge required to begin training for justice at the higher level. We cannot answer, "habituation." The process Aristotle is describing in II.4 is the process we translate with the English word habituation, namely, the active process by which one works to acquire virtue. There is no lower level of parental 'inculcation' where the real work of virtue acquisition happens. We should not confuse Aristotle's famed insistence on a well-brought up audience with a passive conception of how his audience became good. Of course parents and teachers play a significant supporting role¹¹⁰, but it is nonetheless of the utmost importance to Aristotle that the agent is in control of his own virtue-acquisition:

But if without being ignorant a man does the things which will make him unjust, he will be unjust voluntarily. Yet it does not follow that if he wishes he will cease to be unjust and will be just. For neither does the man who is ill become well on those terms—although he may, perhaps, be ill voluntarily, through living incontinently and disobeying his doctors. In that case it was then open to him not to be ill, but not now, when he has thrown away his chance, just as when you have let a stone go it is too late to recover it; but yet it was in your power to throw it, since the moving principle was in you. So, too, to the unjust and to the self-indulgent man it was open at the beginning not to become men of this kind, and so they are such voluntarily; but now that they have become so it is not possible for them not to be so. (1143a12-a21)

In this passage of NE III.5, Aristotle once again takes up the analogy of doctors and their instructions. He deploys his claim from II.4, that we become just through what we do, in the service of the conclusion that we are responsible for ending up just or unjust. Aristotle holds that what makes you just is *what you (voluntarily) do*, not anything that is done to you. It would contravene one of the central argumentative functions of II.4 within the NE as a whole, that of grounding his account of moral responsibility, to push the work of moral education backstage to an earlier period of inculcation.

Jimenez also has trouble making adequate space between trainee-justice and full-justice. If the trainee's action is already just, it is hard to see what the additional conditions add. Indeed, there is, to my mind, an insuperable textual difficulty with Jimenez' claim that the trainee performs the virtuous action for its own sake, since Aristotle specifies this as

¹¹⁰ Though Aristotle seems to think the most important supporting work is done by the laws and the lawgiver (see NE X.9).

one of the three *additional* conditions for acting virtuously, one which, therefore, cannot be met by the trainee. The just man, on Jimenez's picture, boils down to someone who regularly or reliably does what the trainee only sometimes manages. But Aristotle seems to think that there is a bigger gulf between the trainee and the fully virtuous person. He denies that the trainee acts *justly*, even if he does what is just. Aristotle seems to think that the trainee's manner of acting is somehow defective, even if we approach it from only the narrow temporal frame of a single action. In short, Jimenez' interpretation denies the virtue-activation principle.

It might seem that there couldn't be an alternative to the two interpretations I have considered: either the trainee's action doesn't *per se* have ethical value, or it does. As with Strawson's dilemma, I hold that there is an escape from the dichotomy. We can avoid collapsing trainee justice into full justice without placing the former entirely outside the normative sphere. The key is to recognize the normative dependence of trainee-justice on full justice. On the external aspect interpretation, trainee justice can be specified extra-ethically in non-normative terms; on Jimenez' interpretation, it can be specified intra-ethically, without reference to the justice of the virtuous man. For the latter simply *adds* something to what is present already at the level of the trainee. But Aristotle, describing the just and temperate actions of the trainee, says that such actions "are called just and temperate when they are such as the just or the temperate man would do." He seems to suggest that we pick out which trainee actions are just by looking to what the fully just do.

Both of the interpretations I have considered assume that we can grasp the trainee's activity on its own terms. I propose, instead, that the trainee is performing a quasi-virtuous action, an aspirational imitation of what the virtuous man is doing. It is right to use the language of temperance, courage and justice to describe the actions of the trainee, but this is not because they are fully temperate, courageous or just. It is because there is no lower-level normative concept with reference to which we might identify them. The action of the fully virtuous person sets the standard for what the trainee is doing, or rather, trying to do. We understand what the trainee is doing by gauging his distance from the person he is imitating. There is no word to describe the trainee's action but "just," even though what he is doing is not just. The trainee does a just action, namely, a kind of action of which only another person—the fully just person—is a model. The normative structure of trainee action is given by the concept of (full) justice, which it only partly instantiates.

Aristotle's additional conditions on full virtue are simply a way of specifying the fact that the person who does the virtuous thing virtuously has a complete or perfect version of the virtue present in the trainee. When Aristotle contrasts the trainee with the fully virtuous person on the grounds that the latter chooses virtuous action for its own sake¹¹¹, I do not

¹¹¹ Aristotle presents the 'choice' condition and the 'for its own sake' condition as one, but also divides them verbally: "secondly he must choose the acts, and choose them for their own sakes." I will treat them as one condition, but there is some reason to think that Aristotle thinks the trainee cannot act from choice at all. For he does say that choice is an indication that one has the relevant character (1112a). The trainee may belong among those who do not act from choice but act voluntarily, such as children and akratics

think he can be read as suggesting that the trainee does the virtuous action for some end that is fully independent of its virtuousness. For Jimenez is right to press the point that a trainee who acts justly or temperately only for the sake of external gain will never become just or temperate.¹¹² Indeed, it is not clear that these actions would even count, for Aristotle, as just or temperate¹¹³. Where Jimenez, and many others, go wrong is in assuming that Aristotle is contrasting ‘for its own sake’ with ‘from a base motive.’ It is important to remember that Aristotle’s terminology is less technical than what we might expect to find in a modern author: phrases such as ‘for its own sake’ are, for him, highly context-sensitive. For instance in NE X.7 (1177b16-20) he contrasts virtuous action, which is done for the sake of an independent end, with contemplation, which is done for its own sake. Aristotle is drawing the contrast between the trainee and the expert on different grounds, and we should look to the structure of II.4 as a whole to locate them.

The guiding distinction throughout this passage—and the one with reference to which Aristotle hopes to resolve the paradox—is that between *doing virtuous actions* and *being a virtuous person*. I suggest, then, that when Aristotle says that the trainee does not do virtuous actions for the sake of those actions, that is not because he does them for the sake of any base motive, but because he does them *in order to acquire virtue*. Someone cannot do an action for its own sake to the extent that he has one eye on the kind of person doing that action will make him. Aristotle has, here, grasped something of the necessary self-obsession of the young. They are self-obsessed because they are acting for the sake of their future selves; and self-obsession is necessary because it is only way that the right kind of transformation can take place.

Jimenez cites shame to explain how the trainee could be motivated to do what is just without justice. She is right that Aristotle takes shame to play an important role in moral education, but there is no evidence that Aristotle thinks that those who act from shame do what they do for its own sake. Indeed, the natural thing to say is that those who are motivated from shame are thinking not about their action, but rather themselves. Shame is a painful form of self-attention.

The actions of the trainee are not mere *repetitions*. Though he does just things “many times,” his doing of them changes over this time, as he works his way toward his own end of doing them correctly. It is true that *we* cannot classify the trainee without reference to the (fully) virtuous person; but it is even more significant that the trainee cannot understand *himself* without making reference to the (fully) virtuous person. He

(1111b12-14). Aristotle might also hold that trainees cannot deliberate, since we deliberate about what is in our power, i.e., up to us (1112a31), and choice presupposes deliberation (1112a15-16, 1113a3, 1135b9-11). Perhaps Aristotle thinks that just insofar as one is a trainee in a given field, there is nothing that it is up to one to do. Finally, it is worth entertaining the possibility that, for Aristotle, at least in the context of this passage, choosing an action, and choosing it for its own sake, are synonymous expressions. It may be, then, that the attitude of choice is one that one can only have to actions when one does them with an eye to their value, rather than to the changes one hopes they produce in oneself.

¹¹² See the passage from V.8 cited above (fn. 109) for actions that are merely incidentally just, and X.9 (1179b11-16) for how ‘the many’ never become virtuous, since they do virtuous things only for the sake of pleasure or pain or fear (rather than shame).

¹¹³ except in an incidental sense.

gauges his own distance from the person he is trying to be, so as to adjust his activity to his target. In this way, we can restore a kind of explanatory force to the third of Aristotle's three conditions: the trainee does not act from a fixed disposition to justice not only because he fails to be just, but because his disposition is, more specifically, in flux. He has a different kind of character from the vicious (or virtuous, or akratic, or enkratic¹¹⁴) person: an unfixed, changing character.

Aristotle describes for us two ways of doing virtuous actions: the first in the manner of the trainee, and second in the manner of the virtuous person. When he specifies that the virtuous person does what he does virtuously¹¹⁵, he is not adding any additional property, but simply pointing out that the virtuous person is, and sees himself as, doing what the trainee is only trying to do. When Aristotle says that the trainee cannot do the action 'for its own sake' or 'from being the kind of person who would do that action,' he is noting that the very fact that he is *trying* infects and colors the way in which he does just and temperate and courageous actions. The experience of being a trainee is not like the experience of being just, or temperate, or courageous. For trainee-justice, courage and temperance includes a conscious understanding of oneself as not yet being just, courageous, or temperate.

Aristotle is able to say that the virtuous person becomes virtuous by doing what is virtuous, because 'doing what is virtuous' can be taken both in a complete, and a defective sense. A passage from the *Eudemian Ethics*¹¹⁶ helps us differentiate the two senses:

"Excellence too, then, is that sort of condition which is produced by the best movements (tōn aristōn...kinēseōn) in the soul, and from which are produced the soul's best works (ta arista erga) and feelings (kai pathē)...." EE II.1 1220 b29-31

Though excellence is produced by acting excellently, and it gives rise to acting excellently, Aristotle describes the former as a change in the soul (kinēsis) and the latter as the deeds (erga) and feelings (pathē) of the soul. The actions of the virtuous person are actualities—they are deeds whose end lies in themselves, and proper responses (pathē) whose value lies in the fact of their correct attunement. Aristotle thus distinguishes two forms of agency. The trainee's virtuous agency produces changes in his soul. The fully virtuous person's agency, by contrast, allows him to remain in a fixed psychic condition, while actualizing his dispositions to act (erga) and react (pathē) to the world in the correct way. Aristotle describes both the psychic changes (kinēseis) that give rise to virtue, on the one hand, and the active (erga) and passive (pathē) exercises of virtue, on the other, as "best." Nonetheless, the bestness of the former is parasitic on the bestness of the latter.

¹¹⁴ For the claim that the akratic and enkratic have a fixed character, see Coope 2012 pp.151-152.

¹¹⁵ "...done something grammatical and done it grammatically...if the acts that are in accordance with the excellences have themselves a certain character it does not follow that they are done justly or temperately.... is not the man who does these that is just and temperate, but the man who also does them as just and temperate men do them."

¹¹⁶ Though see London, Moral knowledge and the acquisition of virtue, who argues that Aristotle's take on moral education is more intellectualist in the EE than the NE.

The virtue-producing changes in the soul are best among the motions of the soul because they result in the condition that gives rise to the best deeds and feelings. Aristotle's description of the period of time during which the trainee is training as a process of motion or change (*kinēsis*) correctly highlights the fact that it is directed at an end outside itself, rather than being done for its own sake. Unlike an actuality, which is complete at every moment, a process of change (*kinēsis*) is only complete once it has reached its endpoint.

Aristotle has, then, described two variants of acting virtuously. There is virtuous action tout court, and there is the virtuous activity of the learner who is actively seeking to become virtuous in the first sense. The important point is that the virtue of the expert, though it comes later in time, is the source of normativity for the learner. The trainee is trying to see (and think and feel and value) as the fully virtuous man does. If he succeeds, he ends up in a position to choose virtuous actions for their own sakes. It will not be right to see those actions as 'products' of the actions done in his traineeship—even though, had he not done them, he could not act in this way. As a trainee, he was not in a position to choose anything for its own sake, or to act from a fixed character. If he does come, eventually, to act with full virtue, those actions are not (normatively) derivative of his trainee virtue; the reverse is true. His trainee actions were approximations of what was to come later. The key to Aristotle's solution to the paradox with which he opens II.4 is that a temporally prior self can be normatively posterior. The agency of the earlier self exemplifies that special form of psychic change that is *learning*.

If all of this is correct, one would expect nothing to hang on the fact that Aristotle solves the paradox by drawing a distinction between two forms of acting virtuously, as opposed to by distinguishing two forms of having virtue. Given that the solution requires him to differentiate between the agency of the trainee and that of the expert, and that agency manifests in both action and character, it seems he could have done so in the latter sphere instead of the former. Which is to say, he could have explained the ethical difference between the trainee and the expert by distinguishing the former's partial virtue from the latter's complete virtue. It is, then, a point in favor of my interpretation that the latter, alternative way of formulating the difference appears as a response to the same paradox in *Metaphysics* theta 8:

This is why it is thought impossible to be a builder if one has built nothing or a lyre player if one has never played the lyre; for he who learns to play the lyre learns to play it by playing it, and all other learners do similarly. And thence arose the sophistical quibble, that one who does not possess a science will be doing that which is the object of the science; for he who is learning it does not possess it. But since, of that which is coming to be, some part must have come to be, and, of that which, in general, is changing, some part must have changed (this is shown in the treatise on movement), the learner must, it would seem, possess something of the science. But here too, then, it is clear that actuality is in this sense also, viz. in order of generation and of time, prior to potency. (1049b29-1050a3)

In this passage, as in NE II.4, Aristotle aims to make sense of the claim that one, e.g., learns to build by building. Here he articulates the paradox as that of how someone can be learning when he is ignorant. This is just another way of saying that the virtue acquisition principle is in tension with the virtue activation principle: how can someone

be engaged in X-ish activity without possessing the disposition, X? Aristotle's answer here is not to point to two ways of doing X-ish activity, but to point to two ways of possessing the disposition, X. The trainee possesses a partial, attenuated instance of the disposition ("something of the science"), one that allows her to take her own complete possession of it as a target. The virtue of the expert, by contrast, doesn't need to orient itself with reference to an anticipated, completed condition of itself. Thus the virtue of the trainee is 'in motion', whereas that of the expert has reached its *telos* of a completed change. And I have offered aspiration as a way of characterizing the condition of being in ethical motion toward an end.

I hope it is clear that the project of interpreting Aristotle's account of *ethismos* benefits from the aspirational solution to Strawson's dilemma. We found a middle ground between the interpretation which denies the virtue acquisition principle, and the one that denies the virtue activation principle; and we did so by sailing on much the same course as we used to avoid Strawson's Scylla of arbitrariness and his Charybdis of ethical stasis. In denying the virtue acquisition principle, the external aspect interpretation generates a break between the virtuous person and the trainee. The latter seems to arise at random from the former, violating Strawson's continuity requirement. Jimenez' view avoids this disruption of continuity at the cost of denying real change between the two characters, violating the novelty requirement. I argued, in response to Strawson, that genuine self-creation is possible because we can use our dim or inchoate awareness of value to come into contact with the value that, we can then see, has guided the process all along. This is the same possibility to which Aristotle seems to be pointing in his resolution of the paradox of *NE* II.4 and *Metaphysics* theta 8. Aristotle does not say that the resolution is a matter of denying the truth of either the virtue acquisition principle or the virtue activation principle. The aspirational analysis puts us in a better interpretative position by showing us how the two principles can be steered out of the way of each other. It also shows us why it does not, in the end, matter whether one renders the virtue-activation principle consistent with the virtue –acquisition principle by picking out an equivocation in the term "virtue" or one in the term "virtuous action."

II. Aspiration vs. Dialectical activity

In his book *The Retrieval of Ethics* Talbot Brewer describes a set of activities he calls "dialectical":

"Some such activities have a self-unveiling character, in the sense that each successive engagement yields a further stretch of understanding of the goods internal to the activity, hence of what would count as a proper engagement in it. If the activity's constitutive goods are complex and elusive enough, this dialectical process can be reiterated indefinitely, with each successive engagement yielding a clearer grasp of the activity's proper form and preparing the way for a still more adequate and hence more revealing engagement in it." (p.37)

Brewer counts among such activities: parenting a child, having a conversation, being in love, appreciating music. What is characteristic of dialectical activity is that its "value cannot be grasped with perfect lucidity from the outset, but must be progressively clarified via engagement in the activity itself." (p.39) Brewer eventually argues that the

set of dialectical activities is (at least roughly) coextensive with the set of intrinsically valuable activities: "...it seems doubtful that an activity can count as intrinsically valuable for human beings if it is simple and transparent enough that practitioners can arrive at a full and reflectively stable grasp of its highest possibilities and their value." He holds that in order to be able to account for these activities as unified over time—as opposed to merely broken into disconnected stretches—one needs to abandon a number of foundational assumptions in action theory, namely: that desires are propositional attitudes, that they aim to bring about the proposition specified in that content, and that they can be paired with beliefs to produce a rationalizing explanation of action. Brewer calls this set of assumptions "propositionalism," and much of his book is devoted to showing that dialectical activities, and the distinctively "dialectical desires" that motivate one to engage in those activities, call for a non-propositionalist conception of desire.

Brewer's language in describing the agent's grasp of the value and nature of the dialectical activity—that it is partial or unclear or veiled—sounds remarkably like the language I have been using throughout this book to describe aspiration. Likewise, his insistence on understanding dialectical activities as extended over time as opposed to broken into disconnected series of actions may remind the reader of one of my arguments in chapter 2, namely, the argument that proleptic reasons are needed to capture the unity of the aspirational project. I do not propose to assess the main bold theses in Brewer's rich book, but I do want to say why the phenomena he is describing, as he understands them, are neither aspirational nor raise the paradoxes that attend aspirational phenomena; moreover, I do not think that a recognition of the phenomenon of aspiration, taken by itself, requires us to abandon propositionalism. Dialectical activity cannot do the work in my theory that aspiration does, and aspiration could not do the work in his theory that dialectical activity does.

The crucial difference turns on what, precisely, is meant by the idea of a partial or incomplete grasp of value. Let me distinguish two ways in which we might describe some entity as failing to satisfy or instantiate some norm. The first way is by being *imperfect* and the second is by being *inadequate*. Something is imperfect when there is any respect at all in which it does not instantiate the norm. For any given instantiation of a norm, we can often imagine a better or more complete one. Inadequacy is a much more serious flaw. If something is inadequate it is not good enough, in the sense of being in need of some kind of remedy. It is not, as such, a problem that a kitchen table is not perfectly clean; presumably no table has ever been entirely free of dirt or bacteria; but it is a problem if the table is inadequately clean. Brewer doesn't draw this distinction, but a number of features of his view point to the conclusion that when he says the agent does not have a 'full' grasp of the value, he must mean that her grasp is *imperfect*. The aspirant, as I understand her, is someone whose grasp both is, and is known by her to be, *inadequate*.

Let me explain why I ascribe the former view to Brewer, and why the latter plays such an important role in my own theory. Brewer's characterization of some activity or desire as dialectical is not restricted to agents at an early stage of development: he understands the relevant set of activities, which is to say those that are intrinsically valuable, to be

dialectical even for the expert. So even the seasoned practitioner of, e.g. dance, would not find her grasp of the value of dance, and of the activity of dancing, to be, in Brewer's sense, "full." Brewer traces the concept of dialectical desire to the descriptions of the desire for union with God in Gregory of Nyssa, Augustine and Aquinas (as well as later Christian mystics). Infinitude borne of the impossibility of, e.g., "sating one's appetite for God" (p.160) is written into the nature of the activity as dialectical. It would not necessarily follow from this picture that the distance from completion marks the state as imperfect as opposed to inadequate—for one could think that all of human life is marked by a deep failure to be in contact with any true value. But this is not Brewer's view. He spends a chapter of his book explaining how these activities are characteristically pleasant, drawing on Aristotle's conception of pleasure as an *energeia* (activity) complete at every moment. Thus he holds that the failure to have a "full" grasp is compatible with having an adequate or good enough grasp of the value. For Brewer is clear that we do engage in these activities for their own sakes, and that in so doing we find them rewarding sources of meaning as opposed to painful reminders of our defectiveness.

The aspirant, by contrast, is aware of the *defectiveness* of her grasp of some value. She is unable to engage in the relevant activity purely for its own sake, precisely because she does not yet value it in the way that she would have to in order to do so. And because she can see that her grasp of the value is inadequate, she is moved to change herself into someone who has a better grasp. Thus the music teacher's paradoxical demand that the ideal student of her class, as an aspirant, respond to musical value exactly to the extent he is not yet able to. We cashed out this form of responsiveness as a *proleptic* one, in which the agent is moved by a proximate, but acknowledgedly inadequate, representation of value as a stand-in for the real value one hopes to thereby attain. Unlike Brewer's dialectical activity, aspirational activity is not done for its own sake. As we saw in our discussion of Aristotle, the activity of the aspirant characteristically aims at self-change: she is taking a music class in order to become a different kind of person, namely one who appreciates music. Unlike aspiration, dialectical activity is not a mode of self-creation; for the same reason, Strawson's paradox does not arise as a challenge for Brewer¹¹⁷.

Brewer contends that a dialectical desire cannot be expressed in the form of a proposition describing a state of affairs to be brought about. The aspirant's desire can. Qua aspirant, she desires to bring about the state of affairs in which she is a person who can appreciate, e.g., music. This desire suffices to unify her entire aspirational project, from the moment of its inception until she becomes the person in question. Of course if Brewer is right, and the activity of (fully) appreciating music must itself be understood as the object of a non-propositional desire, then the same will be true of aspirational desire. For the analysis of aspirational desire is parasitic on the analysis of the desires of the person in the completed condition. My point is only that the aspiration theory, taken by itself, does

¹¹⁷ Brewer does, in one place, make a glancing reference to a "benign paradox of inquiry" attending "any dialectical activity whose constitutive goods have not been fully mastered." The paradox strikes Brewer as benign precisely because the agent engaging in such an activity is, despite the fact that she has not "fully mastered" the good, nonetheless "sufficiently aware of what she would count as good." (p.47)

not give us reason to resist propositionalism. It is also neutral on the question of whether the completed activities as have, as Brewer supposes, a component of infinite perfectibility.

III. Aspiration vs. Ambition

It has been suggested¹¹⁸ to me that most of the value changes we experience stem not from making ourselves into, e.g., a gourmand or an opera lover, but rather a simple openness to value. Compare a graduate student who selects her courses with the aim of shaping herself into a successful academic, with one driven by interest and passion to take courses that do not necessarily cohere into any pre-arranged program of study. This story could play out in a variety of ways, but it is not merely those of a romantic bent who might anticipate the possibility that the second student might surpass the first academically. The first is, perhaps, working on herself with too heavy a hand. A pair such as these might lead one to caution against “too much aspiration,” lest the aspirant lead herself down too narrow a path on the basis of a blinkered conception of the value she’s pursuing. Are we, sometimes, advised to allow ourselves to develop more naturally, by simply being open to value as it presents itself?

Aspirational pursuits combine the property of being large in scale (we contrasted small and large-scale transformative pursuits, above) with that of being directed at change in oneself. Self-cultivation overlaps with aspiration on the latter front: when we cultivate ourselves, we engage in a pursuit that is self-directed, but small in scale. Because self-cultivation presupposes the priority of the earlier self, it cannot engender major changes, such as reversals in what Ullmann-Margalit calls one’s ‘core values.’ Thus we found that the self-cultivation model of self creation underestimated the possibility for radical, rational, self-transformation. It will, at this point, be useful to contrast aspiration with the kind of pursuit that is large in scale but is not directed at producing a change in the self. These are cases in which the agent aims, usually over many years, to achieve something difficult and important. Nonetheless, the pursuit is not, with respect to value, a learning experience: she enters into it knowing why she is doing what she is doing. As indicated above, I will use the word “ambition” to describe such an agent. An ambitious agent’s behavior is directed at a form of *success* whose value she is fully capable grasping in advance of achieving it. Hence ambition is characteristically directed at those goods—wealth, power, fame—that can be well appreciated even by those who do not have them.

Ambition is often salutary: the researcher who has the ambition of curing some disease, is, if she succeeds, a boon to mankind; likewise the politician who aims to ameliorate social ills or the inventor who discovers ways to make life easier for everyone. But because ambition both consumes much of an agent’s efforts and does not expand his value horizons, it carries with it the danger of trapping him in what may be an impoverished appreciation of value. This danger is lessened if ambition does not entirely usurp the space for aspiration. For it is important to acknowledge that someone might harbor both ambition and aspiration in the same domain, just as someone might

¹¹⁸ By Rachana Kamtekar and Erin Beeghly, to whom I owe the example of the two graduate students.

simultaneously aim to both cultivate himself and to aspire. To the extent that she is ambitious, she seeks to realize values already in her grasp; but she might also, at the same time, aspire to come into better contact with those same values. Plausibly, the great achievers in human history have, for the most part, been both ambitious and aspirational. Nonetheless, it is possible to criticize someone for approaching some pursuit with too much ambition, too little aspiration. This is, I believe, what the example of the two graduate students illustrates.

The “heavy handed” student’s problem, if she has one, is not that she is aspirational as opposed to open-minded, but that she is ambitious as opposed to aspirational. She is trying to succeed instead of trying to learn. Aspiration is not to be contrasted with openness to value; rather, aspiration *is* openness to value. Which is just to say that there is no such thing as generalized openness to value that does not take the form of some concrete pursuit of value. For what would such openness consist in, practically speaking? Being open to the value of poetry involves acting differently than one would in order to be open to the value of yoga¹¹⁹. If openness is to be a genuinely practical attitude, it must be at least somewhat value-specific. There must be a (set of) values that one is, so to speak, trying to value. Or a set that one is trying *not* to value.

For one can aspire to be free from a value rather than to acquire one: my parents’ generation was the first to be faced en masse with the task of revising their sexual norms in the face of encountering homosexuality in their own children, or in friends of their children. They aspired to rid themselves of the values that would dictate attitudes they found to be incompatible with the love they felt for these children. It is, in part, due to their aspirational work that homosexuality has a very different place in our culture today than it did 30 years ago¹²⁰.

In calling for openness, with respect to sexual morality or anything else, we do not ask for people to instantaneously change who they are, nor to do so entirely on their own. The call for openness is a call to try to feel the defect in one’s valuation, either the failure to value something that is valuable, or the mistake of having placed value on something that is not in fact valuable. The call to openness also, characteristically, involves opening oneself to being shaped by others: to listen, to attend, to submit oneself to the judgment of those who have a better grasp than oneself. And this is what the early stages of aspiration look like: having an inkling of interest, being moved to look further, having a sense that one’s preconceptions and assumptions may not be right or that one hasn’t made adequate room for something, having a sense that there are others who have a better grasp, in the domain in question. As I have remarked a number of times, the aspirant relies on others for a distinctive kind of help. While the ambitious person may receive assistance from others in achieving his goal, the aspirant needs others to help her with the project of grasping her goal. The ambitious person may, of course, also be ignorant: she may not know how to acquire some particular means to her end, or how to jump over some hurdle standing between herself and success. The ambitious person’s ignorance is, however, circumscribed by her knowledge: she knows why she wants that means, why

¹¹⁹ See my criticism of Paul in ch1, p.30.

¹²⁰ Thanks to Kristina Gehrman for this nice observation.

she wants to jump that hurdle. The aspirant's ignorance, by contrast, runs all the way down.

IV. Responsibility for Self: An Asymmetrical Account¹²¹

Adults are, to varying degrees, responsible for the kind of people they have become. They play some role in shaping their own interests, passions, idiosyncrasies, and moral sensibilities. Genetics, parenting, and many different kinds of environmental factors also contribute, more in some cases than others. If you become a menace to society after enduring an unusually harsh upbringing, and I do so while having had the benefit of every advantage, I deserve more blame than you do for becoming that kind of person.

These commonsense intuitions point both to the existence of, and to the existence of qualifications on, someone's responsibility for being whatever kind of person she is. We take ourselves to be responsible not only for what we do (or think or feel), but also for being such as to do (or think or feel) those things. The second kind of responsibility is responsibility for the persisting cognitive, emotional, motivational and evaluative condition that I've been calling a person's "self" or "who she is."¹²² We presuppose responsibility for self when we decry wasted talent and laziness: "you had it in you to succeed!"; as well as when we take privilege or adversity into account when considering otherwise similar applicants: "describe the challenges you have faced..."

The aspirational account supports these various intuitions. On the aspirational account, I am responsible for the values I acquire aspirationally. Since aspiration essentially involves the contribution of one's environment, this responsibility will be a matter of degree. Consider an analogy. Two children playing the same, difficult piano piece. They play it well, with both technical accuracy and musical sensitivity. The first has natural talent and a piano-teacher mother who guides every moment of his practicing, for which he further has ample leisure; the other is largely self-taught, less naturally gifted, and must squeeze time for practice between chores and schoolwork. We will be more impressed by the achievement of the second student than the first: in comparison with him, the first child seems to have had a lot of the work done for him by his mother, good genes, privilege. Both children deserve credit for having developed the power to produce such a performance, but the one deserves more than the other.

In the same vein, I will want to say that I am responsible for my value-achievements to the degree that my aspirational work figures in the explanation of my arriving at the endpoint. I become some way by aspiring when my aspiration to be that way figures

¹²¹ Thanks to Steve White for showing me the need to revise my views on blameworthiness, and for helping me do so.

¹²² The concept of character has come under attack by 'situationalists' such as John Doris and Gilbert Harman, to whom Rachana Kamtekar responds with a defense. In accounting for responsibility for self, we don't need to address ourselves to these arguments, for reasons Kieran Setiya has given: situationalists only deny the existence of character in the sense of "unified traits that range across much of what we do, their influence apparent in widely varying situations." (Setiya, p.75) Even if we concede the non-existence of these traits to the situationalist, we can construe the self as composed of dispositions that are more fine-grained, and not necessarily widely shared.

significantly in the causal explanation of my becoming that way. The more it so figures, the more responsible I am. But this is not a complete account; for it describes a merely sufficient condition on your being responsible for your current condition. Or rather, it describes a condition that is both sufficient and necessary for responsibility for being in a *good* condition. I believe we need a different account of responsibility for bad, i.e. blameworthy conditions. I cannot offer a complete account here of why this is the case, or what the account should be. But I find the issue worth addressing in this context, if only in a cursory and promissory way.

It might appear that the account on offer here can easily be extended to bad cases: you are responsible for your good character if you got there by aspiring, and you are responsible for your bad character if you got *there* by aspiring. But there are some reasons to think this may not work. The first problem is presented by the figure of the lazy agent. Jill has, despite ample luck (natural talents, loving parents, etc.), never tried to appreciate any value other than the creature comforts to which her privileged upbringing gave her easy access. Jill's valuationally impoverished condition seems to be her own doing, in the sense that she is responsible for being the (bad) way she is. But she did not *aspire* to be the way she is. Indeed, her problem is that she did not aspire to be any way at all: Jill's problem is complacency.

I suggest that we accommodate Jill's responsibility for her condition by invoking the possibility that someone can culpably fail to aspire. She can be blamed, in effect, for her very self-satisfaction. We can use environmental factors to distinguish (more) culpable failures to aspire from non (or less) culpable ones in much the way that we do in cases of credit, *mutatis mutandis*. Having one's route to some positive condition facilitated by privilege, genes, and good luck makes one less responsible for success than the one who had to struggle. In the case of a failure to, e.g., develop one's talents, we can likewise say that an agent is less culpable for his failure to develop himself if major obstacles stood in the way. Jill would be the contrast case, an agent whose failures to aspire are not due to her environment, but are directly attributable to her. Thus we can open up, on the side of responsibility for bad character, an avenue of what we might call omissive culpability. This already suggests an asymmetry between the bad case and the good one: on the good side, there is only commissive responsibility, whereas on the bad side, it seems there is both commissive and omissive responsibility. And yet I think that the asymmetry runs even deeper.

For how exactly are we to understand commissive responsibility for being bad? Given that aspiration is an intentional state, no one, except perhaps the devil¹²³, aspires to be bad. Suppose we have before us an evil, cruel, heartless gangster. He denies being, or having aspired to be, any of these things. He insists that both is and aspired to be nothing other than a great leader in his community. Can we hold him responsible for the condition that he ends up in (public scourge), if he ended up being that way as a result of having aspired to be a *different* way (public servant)? To do so would be to introduce an opaque form of self-creation. We will not be able to understand commissive

¹²³ And this case is notoriously hard to understand. cf. Anselm *Casus Diabli*, on which see Tomas Ekenburg 2015.

responsibility for being bad on the model of commissive responsibility for being good. It matters, for our conception of the process of aspiration as an agential one, that the goodness one achieves is the very goodness one was (proleptically) after. Otherwise, one is not creating the very self that one becomes.

We cannot say that the dictator is responsible for becoming cruel because he became cruel while aspiring (to be great). For there are cases that fit this formula where we will not want to hold the agent culpable. For instance, suppose it is a result of my successful aspirations to become a great pianist that everyone in my music school comes to hate me for my superiority to them. I become hated by all as a result of aspiring, but I did not aspire to be hated by all. Nor does it seem to be the case that I am responsible for being hated by all. We want to distinguish the evil dictator from someone whose bad condition is a mere side effect of her aspiration. There are reasons, then, to think that someone is only to count as having aspired to some condition she is in if being in that condition figured—however obliquely or opaquely—in the intentional content of her aspirational mental states (desires, beliefs, feelings etc.).

If this is so, then in virtue of what is someone responsible for becoming an evil person? I want to suggest, in the face of the difficulty of constructing a commissive account of such responsibility, that the error must, like lazy Jill's, be one of omission. The evil person could have, and did not, aspire to become a good person. He had enough contact with the beginnings of goodness through his early childhood that he had the materials from which such an aspiration could spring. I am not, of course, saying that his *actions* constitute omissions. But it is possible to be responsible for being the kind of person who would *commit* those kinds of actions in virtue of having culpably *omitted* to aspire to be a better kind of person.

If this account of responsibility for self is correct, the asymmetry introduced by aspirational accounts of moral responsibility is a deep one indeed. We are responsible for good ways that we are to the degree that we aspired to be in those ways, and bad ones to the degree that we culpably failed to aspire to be better. The analysis of the bad cases would, on this account, be conceptually derivative of the good ones.

V. The Aspiring Gangster?

But is it true that no one aspires to be bad? Consider a gangster in training. It looks as though “gangster” is a description under which someone is blameworthy, and at the same time a description of a way someone might aspire to be. It is imaginable that someone might try to become a gangster, under that very description. Could someone *aspire* to become a gangster? Recall our two graduate students. Let us exaggerate the differences between them so that the one is maximally ambitious, the other maximally aspirational: the one chooses her courses entirely with the aim of impressing the people whose letters will carry weight; she writes on the most popular topic; she fills her dissertation with the kinds of phrases and points that she imagines will make her sound impressive and right thinking, etc. The other student chooses advisors because she admires their work and thinks they have something to teach her; she picks a dissertation topic from interest and passion, etc.

These students may be caricatures, but the phenomenon I am caricaturing is very familiar to anyone who has been a teacher. Turning ambition into aspiration is one of the job descriptions of any teacher, but it belongs especially to those of us who, as philosophers, carry the Socratic torch. Socrates may not have had great success with, e.g., Alcibiades, Thrasymachus, and Callicles, but there is no question that he was *trying* to transform their ambitious political strivings into an aspiration to virtue. Aspirants make good learners because, quite simply, they take themselves to have something to learn. We saw that *ambition* was marked precisely by value-stasis. The ambitious person, qua ambitious, is engaged in getting what he wants, as opposed to learning what he wants.

Now return to the gangster in training. This is a popular figure in literature and (especially) movies. All of us have encountered representations of someone working to become a gangster. In order to show that any one of them was an aspiring gangster, we would have to imagine what differences we would introduce into the story in order to retell it as one of mere ambition. And it seems to me we cannot do this. Unlike with the two graduate students, we cannot imagine two young gangsters who are distinguished by the question of whether they aspire or merely have the ambition to become gangsters. We might imagine approving of one such gangster more than another, but this is only because we take him to aspire not to being a gangster, but rather, to become someone who, e.g. helps his family and gets respect. But aspiring to being that kind of person is aspiring to something *good*.

It might well be that, at the early stages, the gangster-in-training doesn't understand what it is to be a gangster: he identifies it with the list of good traits named above. He says he wants to be "just like my father" but that is because he doesn't know what his father does. But what does he do when he learns the true meaning of, e.g., "take care of" so-and-so? If he rejects the life with horror, we can say that he *had* aspired to be good, and learned that the path he was taking toward that aspiration was mistaken; if he continues to work to be a gangster, his state was never anything but ambition. As an illustration of such a case, consider Henry Hill, the protagonist of *Goodfellas*. He describes his initial attraction to mob life:

"As far back as I can remember, I always wanted to be a gangster. To me, being a gangster was better than being President of the United States. Even before I first wandered into the cabstand for an after-school job, I knew I wanted to be a part of them. It was there that I knew that I belonged. To me, it meant being somebody in a neighborhood that was full of nobodies. They weren't like anybody else. I mean, they did whatever they wanted. They double-parked in front of a hydrant and nobody ever gave them a ticket. In the summer when they played cards all night, nobody ever called the cops."

Contrast a speech that might have been given by a reminiscing writer. I'll call him, "Stephen," after Stephen King:

"As far back as I can remember, I always wanted to be a writer. To me, being a writer was better than being President of the United States. Even before I first wandered into the newspaper office for an after-school job, I knew I wanted to be a part of them. It was there that I knew that I belonged. To me, it meant being somebody in a neighborhood that was full of nobodies. They

weren't like anybody else. I mean, they did whatever they wanted. They didn't own TVs or care what other people thought about them. They read during meals and ignored social conventions; they were rude and they paid no penalty for it."¹²⁴

Both speakers use aspirational language, describing a desire to “be somebody” and to be free from social norms or constraints. They both express admiration for a group they take to be powerful and would like to be part of. Nonetheless, we interpret them quite differently. The writer is plausibly describing an early stage of a learning process, a stage in which he exhibited an inkling of what was to come. The gangster’s speech does not point beyond itself. Instead, it indicates to us that he is trapped in a fantasy. There is no analogue for the speech King makes advising young writers: “Writing isn’t about making money, getting famous, getting dates, getting laid or making friends.” (p.270) For becoming a gangster *is* simply about those things, and nothing more. Perhaps there are gangsters who would insist, at the end of their lives, that there was something they were after all along: respect. But we would take a reductive attitude to the phenomenon the gangster calls, “respect,” denying that it has real value and thereby analyzing it in terms of his associates’ attempts to stay on his good side for their own self-interest. The gangster isn’t honored, he is simply feared.

When young Henry and young Stephen look at the people they admire, they are both attracted by a kind of power. But Stephen’s grasp of that power is aspirational: in pursuing it, he is also at the same time pursuing a value of which it is an intimation. He is reaching beyond himself, proleptically, in the manner of the aspirant. Henry’s reason has no distal face. He can’t be aspiring because there is no “there” where he is trying to go. Suppose there is a gangster otherwise like Henry, but who says that “all I’m after, as a gangster, is money, power and women.” He would not be a different kind of person from Henry. The fact that such a person describes himself in cynical terms, whereas Henry employs more romantic language, has no ethical significance. It doesn’t *matter* that Henry talks in terms of wanting to be a better person; the defective nature of his target as prevents him from counting as the aspirational counterpart to the cynical gangster in training.

If it is true that seeing someone as an aspirant presupposes some appreciation for the value toward which she aspires, then we find in the phenomenon of aspiration a tight connection between our adopting a certain form of explanation for some bit of behavior, on the one hand, and our approval of that behavior, on the other. We find an argument for a view of this kind in the opening discussion of Augustine’s *On Free Choice of the Will*. Evodius asks Augustine “From whom did we learn to sin?” Augustine answers that we cannot “learn” to sin, because learning is the process of acquiring knowledge: “since learning is good, and the word ‘learning’ (disciplina) is correctly applied only when we

¹²⁴ Inspired by passages from King’s *On Writing*, such as “Reading at meals is considered rude in polite society, but if you expect to succeed as a writer, rudeness should be the second to least of your concerns. The least of all should be polite society and what it expects. If you intend to write as truthfully as you can, your days as a member of polite society are numbered, anyway.” “Life isn’t a support system for art. It’s the other way around.” “If you’re just starting out as a writer, you could do worse than strip your television’s electric plug-wire, wrap a spike around it, and then stick it back into the wall. See what blows, and how far. Just an idea.” [REF]

come to know something (*nisi a discendo non dicta est*), we simply cannot come to know evil things....It follows that doing evil is nothing but turning away from learning.” (p.2). One might worry that Augustine’s point is a merely semantic observation about the word “learn.” Perhaps we do not use the word learn to include cases of learning what is evil, but couldn’t we? Are we making a principled distinction when we refuse to call the person who *takes herself* to be arriving at an improved cognitive condition “a learner”? Evodius presses Augustine on this question when he suggests that there might be “two sorts of learning: one by means of which we learn to do right, and another by means of which we learn to do evil.” Evodius can be heard as asking after the justification for drawing our linguistic (and conceptual) boundaries as we do. Augustine responds by asking him:

Augustine: “Do you at least consider understanding good?”

Evodius: “Certainly. I consider it so good that I cannot see how any human trait could be better. And I would in no way say that understanding can be bad.”

Augustine: “When someone is being taught but does not understand, would you say that he has learned?”

Evodius: “Of course not.”

Augustine: “Well, then, if all understanding is good, and no one who does not understand learns, then everyone who learns is doing good.” (p.1-3)

Augustine argues by way of the concept of understanding, but it is not immediately obvious why Evodius couldn’t simply repeat his move, and insist on two kinds of understanding. I want to make a suggestion as to how Augustine leverages Evodius’ denial of the title “learner” to one person, namely the one who doesn’t understand what he’s taught, into a denial of that title to another person, namely the one who “learns” evil. If we were to call the person who “learned” evil things a learner, we would have to do so on the basis that he appeared (to himself) to learn: he took himself to be arriving at an improved cognitive condition. But the person who is being taught appears (to others) to learn. He sits at the foot of the teacher as she expounds truths. And yet that appearance is no basis for calling him a learner. Why not? Because someone who is to learn must *really* be learning and not only be appearing to learn. This justification speaks to the work we want our concept “learn” to do—we use it to pick it out as a concept of making a certain kind of actual progress toward grasping the way things are. The person who is merely being taught fails to make any such progress—but the same is true of the one “learning” evils.

Augustine’s argument aims to expose the internal connection between learning and progress toward enlightenment. I have, in the same vein, described an internal connection between aspiration and genuine value-advance. The two arguments are not, of course, merely analogous. For I have been developing a conception of aspiration as a form of (practical) learning. Insofar as Henry strikes us as having less in common with Stephen than with the cynical gangster, we manifest a concern with aspiration as a form of learning. We could, of course, take a cue from Evodius and carve things up differently. We could speak of appearing to oneself to aspire, a condition shared by Henry and Stephen. Like Augustine, I think it is no accident that we are not antecedently

inclined to make use of this category. Aspiration is what excites our interest, because aspiration picks out something of ethical significance. Whether or not someone appears to himself to aspire is not an ethically significant fact about her; by contrast, whether or not someone aspires matters directly to our assessment and treatment of them, and to the quality of the life they lead. Aspiration is a useful ethical category in a way that appearing to oneself to aspire is not. Let me end by considering an illustration of this fact.

VI. Responsibility for Character: A Test Case

On Dec. 11, 2013, Ryan Loskarn, chief of Staff to Tennessee Senator Lamar Alexander, was arrested on charges of possessing and distributing extremely graphic and violent child pornography. His arrest was met with shock, dismay, disapproval and an effort on behalf of his former associates to distance themselves from him as much and as quickly as possible. In an instant, Loskarn went from an up-and-coming political player to being branded with that special kind of evil mark that we reserve for crimes we don't ever want to think about. A month later, Loskarn committed suicide. After his death, a letter was found in which he revealed that he had been sexually abused as a child, at ages 5 and 9. The letter does not appear to be a suicide note; rather, it seems that he decided to commit suicide sometime after writing it.

Loskarn's letter was met with a wave of empathy, such as we find in expressed in Ruth Marcus' *Washington Post* editorial, "The Tragedy of Ryan Loskarn." Marcus writes, "Some people do terrible things because they are purely evil, others because they are terribly damaged." Seeing Loskarn as a victim of abuse re-framed for us what it meant that he was, as he described himself, an abuser of young children. Details of his troubled past seem to mitigate his responsibility for ending up as he did, since they suggest that, to some degree, something other than Loskarn himself made him into the person he became.

When philosophers discuss examples of agents like Loskarn, they are usually interested in the question of how diminished responsibility for character might detract from someone's responsibility for the actions he performs from that character¹²⁵. That is not the direction of reasoning suggested by Loskarn himself in his letter¹²⁶:

"I understand that some people – maybe most – will view this as a contrived story designed to find some defense for defenseless behavior. That it's an excuse....But I'm sharing this with you because it is the truth, not an excuse. And I believe it played a role in my story."

Loskarn ends his letter with an apology "to the children in the images": "I should have known better. I perpetuated your abuse and that will be a burden on my soul for the rest of my life."

¹²⁵ For a canonical example, see Gary Watson's 1987 discussion of the vicious murderer/victim of brutal child abuse Robert Alton Harris.

¹²⁶ Parts of Loskarn's letter are cited in Marcus' article, and it can be found in its entirety at <http://www.jesseryanloskarnslastmessage.com/>

Loskarn apologizes to the children because, as he says, he sees himself as responsible for having perpetuated their abuse. He does not seek to be exculpated for responsibility for, as he puts it elsewhere in the letter, “my crime.” He offers the details of his abuse because he believes “it played a role in my story.” It seems likely that the role such an event would have to play in the self-told story of one convicted for such an offense is exculpatory. Yet Loskarn does not show any signs of seeking exculpation for what he did (or thought or felt). Recall Marcus’ description of Loskarn as someone who “did terrible things” but only because he is “terribly damaged.” The line plays on the word “terrible,” first invoking a sentiment of righteous indignation at the immorality of his (presumably voluntary) evil actions, and then shifts to an empathetic sadness in response to Loskarn’s status as a victim with respect to who he *is*. She seems to want to exculpate Loskarn for having become the kind of person that he is, despite holding him responsible for what he actually did. In what follows, I will take for granted both the fact that we can be (to some degree) responsible for our characters, and that Loskarn was, and was reasonable in, citing his abuse as grounds for exculpation in respect of responsibility for self as opposed to responsibility for action¹²⁷.

It is unclear where this leaves us with respect to his responsibility for what he did. One might think that, to the extent that Loskarn is not responsible for becoming a pedophile, he is to that extent not responsible for consuming and distributing child pornography. For one might think that any responsibility one has for the actions one performs is derived from one’s responsibility for acquiring the character that would dispose one to perform those actions. This is precisely the argument made by Galen Strawson (1994) in the paper¹²⁸ in which he offers the dilemma discussed earlier, calling into question the possibility of self-creation. He takes responsibility for action to rest on responsibility for character, and he takes the latter to (impossibly) call for self creation. But we should, at this point, be wary of assuming, simply on the basis of the temporal priority of issues of responsibility for self, that it serves as normative ground for responsibility for action¹²⁹. I do not, therefore, take a mitigation in Loskarn’s responsibility for character to entail a

¹²⁷ Responsibility for self includes responsibility for cognitive, emotional and practical dispositions, whereas the action-responsibility that I forbear from discussing covers all, and not merely agential, activations of those dispositions. Angela Smith, developing a strand of thought introduced by Robert Adams (“Involuntary Sins”) and Tim Scanlon (WWO, pp. 283–85) argues that my responsibility for what I think or (don’t) feel cannot be understood in terms of my having *chosen* the attitude in question. Instead, she thinks we ought to construe such cases of responsibility in terms of the subject’s rational answerability for the thought or (lack of) feeling; and that we are responsible when those thoughts and feelings are traceable to our own evaluations. The phrase “responsibility for attitudes” obscures an important distinction between the *disposition* to, e.g., harbor racist beliefs or be jealous of one’s sister and the *activation of such a disposition* in the form of thinking a racist thought or feeling a pang of jealousy. Both can plausibly go under the name of an “attitude.” Her reasons-account is persuasive as an account of cases where we blame someone for what he said, thought, felt, did or didn’t do on a given occasion, but not as an account of blame for *being such as to* do or feel or think in that way. I do not think this is a flaw in her account, since I do not think she is concerned to discuss moral responsibility for dispositions.

¹²⁸ As well as in his 1986 book.

¹²⁹ See also the substantial body of work that Strawson’s argument has generated (e.g. Fischer 2006 (pp.112-118), Clarke 2003 (pp.170-6) and 2005, Mele 1995 (pp.221-230)) much of it addressed to denying or qualifying precisely this premise in Strawson’s argument.

corresponding mitigation in his responsibility for what he did. Nor, however, do I assume to the contrary. I propose simply to set aside the question of action in order to focus on responsibility for self¹³⁰.

Supposing the reader shares my intuition that the revelations in Loskarn's letter mitigate his responsibility for self, let us compare a few ways of accounting for this fact. I begin with the self-endorsement model discussed above (ch. 5, part I). Recall that on this account, we distinguish those features of an agent for which she is responsible from those for which she is not by looking to her own self-evaluations. We are familiar with the phenomenon of agents such as akratics or addicts who evaluatively reject some part or aspect of their motivational makeup. It might appear that a view on which evaluative self-endorsement is necessary for responsibility is well placed to exculpate Loskarn, given his expressions of self-revulsion. I think that a closer reading of his letter suggests, however, that we cannot explain Loskarn's diminished responsibility by pointing to the negative outcome of a self-reflective procedure:

"I found myself drawn to videos that matched my own childhood abuse. It's painful and humiliating to admit to myself, let alone the whole world, but I pictured myself as a child in the image or video. The more an image mirrored some element of my memories and took me back, the more I felt a connection.

This is my deepest, darkest secret."

"In my mind I instigated and enjoyed the abuse – even as a five and nine year old – no matter the age difference.... By my late teens I reached a sort of mental equilibrium on the matter. I couldn't stop the images from appearing altogether, but I generally controlled when they appeared... As an adult I thought I was a tougher man because of the experience; that I was mentally stronger and less emotional than most. I told myself that I was superior to other people because I had dealt with this thing on my own. Those I worked with on the Hill would likely describe me as a controlled, independent, and rational person who could analyze a situation with little or no emotion. That's how I viewed myself. In retrospect, the qualities that helped me succeed on Capitol Hill were probably developed partly as a result of the abuse and how it shaped me."

In many ways, Loskarn identifies both with his abuse ("in my mind I instigated and enjoyed" it) and with the person his abuse made him: "controlled, independent, and rational." He writes, speaking in the present tense, that "in my heart I still struggle to see my five-year-old self as a victim." Loskarn says he 'felt a connection' to child pornography, and he does not present himself as safely alienated from this connection. He does not describe his desire as a compulsion which he can evaluatively condemn as in any sense 'external.' He is unlike Frankfurt's unwilling addict, whose will retreats to the safety of second-order rectitude. Loskarn cannot seem to make a clean break from the

¹³⁰ In drawing this distinction, I do not intend to take a stand on the debate between D. Shoemaker and A. Smith as to whether moral responsibility comes in multiple species ("answerability, accountability and attributability") or can be analyzed using a single underlying evaluative notion. For even if Smith were right that there is single underlying notion, it might manifest in different ways depending on the differences in the object—self vs. action—for which the person is to be held responsible.

negative features of himself—and this is because those negative aspects run very deep. The self-endorsement theory cannot exculpate Loskarn, because he does not succeed in performing the relevant self-alienating task.

The self-cultivation theory does not fare better. Loskarn describes his present self as a product of the “self-control” he learned to exercise to cope with his abuse. In retrospect, he can see that he was “shaped” by his abuse, but at the time he understood himself as cultivating virtues of independence, rationality and control. When describes “the mental equilibrium I had created to deal with my past,” he understands his pedophilia as something he *did* to himself as a way of confining the problematic parts of himself. It looks as though there is some basis for saying that Loskarn *did* cultivate the condition he ended up in.

At this point, we might wonder whether *any* attempt to ground responsibility for self in a kind of agency we have in relation to our selves will be able to account for the case of Loskarn. This is just the point made by Susan Wolf (1987), who assimilates what I have called the self-endorsement view and the self-cultivation view under the label, “the deep self-view,” according to which:

“our wills are not just psychological states *in* us, but expressions of characters that come *from* us, or that at any rate are acknowledged and affirmed *by* us. For Frankfurt, this means that our wills must be ruled by our second-order desires’ for Watson, that our wills must be governable by our system of values; for Taylor, that our wills must issue from selves that are subject to self-assessment and redefinition in terms of a vocabulary of worth.” (p.49)

“the deep-self view... allows us to distinguish cases in which desires are determined by forces foreign to oneself from desires which are determined *by* one’s self—by one’s “real,” or second-order desiring, or valuing, or deep-self , that is.” (p.51)

My “deep self” is in charge of governing or determining what forms of motivation will count as true expressions of my agency. When some desire can be traced—either as endorsed, or as cultivated—to this deep self, then it represents a desire that is truly ‘mine.’ Wolf suspects that all such views will inevitably crumble under regress. “No matter how many levels of self we posit, there will still, in any individual case, be a last level—a deepest self about whom the question, “What governs it?” will arise, as problematic as ever.” She worries that the regress can only be ended with some implausible picture of a “prime mover unmoved, whose deepest self is itself neither randomly nor externally determined, but is rather determined *by* itself—which is, in order words, self-created.” (p.52) She takes the metaphysical difficulties attending the deep-self view to expose the paradoxical quality of the very idea of self-creation. I have argued that the regress that Wolf takes to threaten the very idea of self-creation in fact threatens only one account of self-creation: the account in which the creator self rules over, determines, controls or governs the self it creates. Self-creation needn’t be cashed out in terms of the normative priority of the creator self over the created self. But I want to set aside Wolf’s metaphysical scruples with the very idea of self-creation, and discuss

an objection she raises to the idea that, even if we *could* make sense of such self-creation, it would not suffice for moral responsibility.

She describes a case in which a sadistic dictator, Jo, has indoctrinated his son JoJo into his evil values so well that JoJo is fully on board with Jo's inhumane political agenda:

"JoJo is given a special education and is allowed to accompany his father and observe his daily routine. In light of this treatment, it is not surprising that little JoJo takes his father as a role model and develops values very much like Dad's. As an adult, he does many of the same sorts of things his father did, including sending people to prison or death or to torture chambers on the basis of whim. He is not *coerced* to do these things, he acts according to his own desires. Moreover, these are desires he wholly *wants* to have. When he steps back and asks, "Do I really want to be this sort of person?" his answer is resoundingly, "Yes," for this way of life expresses a crazy sort of power that forms part of his deepest ideal."

Wolf introduces JoJo to show how even someone who acts from his 'deep self'—motivated by values that he endorses, as opposed to ones from which he is alienated—can lack (full) moral responsibility. Comparison between JoJo and Loskarn serves to support Wolf's general intuition that there is a condition on moral responsibility for one's motivational condition that is tangential to questions of whether one endorses or has cultivated it. JoJo enthusiastically avows the values from which his actions spring; plausibly, he has also done work to mold himself into the image of his father, cultivating the character he now has. Loskarn, in quite a different way, identifies with his abuse, internalizes evaluative attitudes that are products of it, and cultivates the character he ends up with. Since she takes the case of JoJo to show that exercises of agency over oneself fail to serve as ground for responsibility, Wolf grounds our mitigation of JoJo's responsibility in a substantive fact about his ethical point of view: his desire for power, and his ethical ideal more generally, is "crazy". JoJo is not responsible for his character because his ethical point of view is insane.

To say that JoJo is insane is, according to Wolf, to say that he lacks "the minimally sufficient ability cognitively and normatively to recognize and appreciate the world for what it is." (p.56) JoJo's ethical point of view is divorced from reality, and he has no prospect of correcting it. I think there is something right about Wolf's approach, in that it suggests that if we want to understand what is wrong with JoJo and Loskarn, we cannot avoid making mention of the substantive fact that what they value is not valuable. They are mistaken or incorrect in their valuational approach. The aspiration approach likewise rules out the possibility that Loskarn's disposition to watch child pornography was his own creation, precisely on the grounds that watching child pornography is bad. There is no value that one's cultivation of such a disposition would be getting at. But I think the invocation of such a substantive condition is not incompatible with claiming that the condition in question is a fact about that person's self-directed agency. Let me motivate why we might want to do this by criticizing Wolf's sanity account.

First, it is not clear to me that Loskarn's values are rightly described as insane in Wolf's sense. Given his inclination to keep his viewing of pornography secret, it seems that Loskarn was not *completely* "unable to understand and appreciate that an action fell outside acceptable bounds." (p.61) Even if there is some sense in which he did not understand the depth of the evil he was committing, it seems wrong to say he was

constitutionally unable to understand this. For by the time he writes the note, he has come face to face with the horror of what he has done. I am not even sure that JoJo qualifies as insane in Wolf's sense. Was JoJo *unable* to appreciate the moral law? Is there *no* set of circumstances we can imagine on which he comes to acknowledge the error of his ways? Since JoJo is fictional, we could simply postulate that there is literally not a single possible world in which some kind of moral awakening occurs during his adult life. But then the JoJo example will be less than useful in our attempt to analyze the ordinary human phenomena of responsibility. We will want to know how actual people, for whom change tends to be (in a broad sense) possible, if unlikely, can be exculpated for characters for which they are not responsible.

The second problem with Wolf's analysis is that it does not draw sufficiently on the historical conditions under which JoJo acquired his ethical perspective. It is not the fact that his views are bad or insane, but the manner in which those views were developed, that seems to have exculpatory bearing. As it stands, Wolf's account cannot distinguish between JoJo and Jo, whose values are just as insane as his son's. Wolf never directs us to mitigate Jo's responsibility; indeed, she herself seems to tie her exculpation of JoJo to his upbringing: "it is unclear whether anyone with a childhood such as his could have developed into anything but the twisted and perverse sort of person that he has become." (p.54) We should not conflate the claim that JoJo couldn't, given his upbringing, have become anyone else, with the claim that he couldn't, having become that person, come to later appreciate morality. Even the former claim seems too strong—perhaps someone with a less accommodating temperament than JoJo would have more seriously rebelled against his father. It certainly does not seem right to insist that, given his childhood trauma, Laskar *had* to develop a child pornography habit. But we seem to be on the right track if we are thinking about the process by which JoJo and Laskar, by contrast with Jo, came to be the way they are¹³¹.

On the aspiration account, we must refer to this path, and to the nature of the agency contained within it, in order to understand whether the person is responsible for making himself who he became. The aspirational account is asymmetrical in that responsibility for good character will be a matter of having made oneself into the person who has that character, whereas responsibility for bad character cannot be understood in a parallel way. The person who ends up with bad values never aspired his way there. We must, then, distinguish culpable from non- (or less) culpable failure to aspire. On what grounds do we judge that a person is less than fully responsible for having failed to aspire to correct his bad values, or acquire good ones?

In our discussion of proleptic reasons, we saw that they characteristically manifest in a social form. Aware that they cannot fully get their target into view, aspirants draw on the

¹³¹Wolf might deny that Jo and JoJo are equally insane. For she could claim that Jo, unlike JoJo, has the *ability* to normatively recognize and appreciate the world, though he does not exercise it (p.62). But she could only do this by insisting on a systematic connection between, on the one hand, the process by which one arrives at (or has inculcated in one) a normative outlook and, on the other, the manner in which one holds the outlook thus arrived at. The aspirational theory of self-creation offers up the philosophical materials with which to underwrite such a connection.

help of teachers, mentors, parents, coaches, therapists, friends, and fellow aspirants. They interact with, are guided by, and also imitate and pretend to be like the more (or in some cases, just differently) the people around them who have a more developed form of the value. They rely on their environment not only to supply them with the minimal grasp of value with which they begin, but throughout the process of growing that initial seed into a complete, established form of valuing. Because the defects in her form of valuing do not escape her notice, the aspirant is motivated to subject herself to the assessment of others. Unlike either self-endorsement or self-cultivation, aspiration is not a theory of solitary self-creation. And this allows the aspirational theorist to cite the fact that someone was cut off from the relevant kind of help in her explanation as to why such a person should not be blamed for her failure to achieve the relevant goal. The aspirational theory places great emphasis on the following common ground between Loskarn and JoJo: they are both relatively isolated from the system of value we would have them learn. In JoJo's case this is a product of the fact that he has been brainwashed by a powerful tyrant. In Loskarn's case, the cause is much more mundane:

"As a child I didn't understand what had happened at the time of the abuse. I did know that I must not tell anyone, ever."

"I always worried someone might look at me and know, so I paid close attention to others for any sign they might have figured it out. No one ever did."

Loskarn did not seek out help because he could not bear anyone to know what had happened to him. He kept himself apart from others. He tried to keep the abuse even from himself, rewriting his memories ("...I instigated and enjoyed the abuse") even as he could see that he was falsifying them. He says, "To those who choose to sever all ties with me, I don't blame you. No one wants to think or talk about this subject matter. All I can say is: I understand and I'm sorry." He understands why people would want to be dissociated from him, because he wants to be dissociated from himself, describing himself as "disgusting and horrible." He expects no sympathy, no empathy, no help, nothing but rejection from his community.

Aspirants do the work of becoming new people, but they cannot typically do this work alone. Loskarn eventually does come to want to change his own ways of thinking and dealing with the abuse, but only as a result of a chance occasion to share his past:

"In my life, I had only ever mentioned the abuse to three friends, and then fleetingly so. I never spoke to a mental health professional about this or any other matter until I was in the D.C. jail. I talked with a counselor there about my crime and the horrible hurt I had caused so many people. I didn't talk to him about my past. I didn't think it mattered because I intended to kill myself as soon as possible.

The session ended and I left to be taken to a cell. Before I'd gone far, the counselor called me back. He said there was something he couldn't put his finger on and he wanted to talk some more. And then he just stopped and looked at me, not saying a word. He was the first person in my life who I think had figured it out. And he was the first person I ever spoke to in any detail about those memories."

It is only *after* speaking to this counselor that Loskarn is moved to write the letter in which he speaks of having endeavored to "begin the process of trying to sort this out and

fix myself.” Until that point, his approach was that of making do with the self he had, rather than acquire a new one. He had recourse only to self-management, controlling limiting and repressing, but never *changing* the “broken” parts of himself.

Loskarn’s experience of abuse is certainly sad, but what is really tragic is his subsequent imprisonment, by shame and secrecy, in his own valuational error. Because Loskarn was a real person, it is easy to have an emotional reaction to reading his letter; and I think the pang of sadness that we feel when we imagine his loneliness contains some intelligence. We wish we could reach out to the child Loskarn, the adolescent Loskarn, the young adult Loskarn and draw him out from his very private suffering into the light of our shared space. He is not responsible for failing to aspire out of his condition because his feeling of shame cut him off from any of the resources such aspiration would have had to draw on.

The aspirational theory is well-placed to explain why those who have suffered from unusually harsh conditions in their upbringing are less responsible for failing to create themselves as good people, and as valuers of the good. We needn’t imagine that an upbringing such as JoJo’s, or a traumatic experience such as Loskarn’s, is incompatible with aspirational success. All we need for exculpation is that such a success, though imaginable, calls for an aspirational task at which most of us would have failed. This is, I think, one payoff of the aspirational account of self-creation in the domain of moral responsibility: because it shows us how the claim that people create themselves is consistent with the admission that they need help to do so, it offers us the correct interpretation of cases such as Loskarn and JoJo.

* Conclusion: Recognizing Aspiration

If one can, contra Nietzsche, “pull oneself up into existence by the hair, out of the swamps of nothingness,” it is only because one is not alone in the swamp. We saw that Aristotle’s familiar emphasis on the importance of parents, teachers and lawmakers for moral education did not prevent him from having an active conception of the process of virtue acquisition. Aristotle thinks that becoming virtuous is something that you do, even if the assistance you get from your city is crucial to the success of your project. In our discussion of Loskarn, we found ourselves making the same point with the opposite emphasis: from the fact that people can create themselves, it does not follow that they have no need of parents, teachers, coaches, and, more generally, helpers. Aspirants are characteristically needy people, since (they see that) their own conceptions of value are insufficient. Their group of supporters includes, but is not limited to, people who may have a better grip than they on the particular values and reasons of the person they want to become. They also rely more generally on the kindness or empathy or material assistance of those who love them, since they tend to make mistakes; to need help; to not know exactly what they are doing.

The vulnerability of aspirants puts them in a distinctive ethical category of those whose treatment by others is a matter of especially deep ethical significance. This, I take it, is part of what explains our profound shock and anger in the face of cases of child abuse. I believe that we have an especially stringent set of ethical intuitions that govern the ways in which it is appropriate to behave toward aspirants. We sense a special responsibility when we interact with our children, our students, our advisees, those who look up to us, and in some cases also siblings and friends. Our feeling that carefulness, sensitivity, tact and empathy are called-for in these interactions betrays an awareness that these people need our help to become who they are. If we do not do enough, they may never get there; and if we do too much, we get in their way.

The modifications I have suggested to the theory of practical rationality, moral psychology and moral responsibility are designed to support the ethical intuitions that we make use of in these interactions. The intuitions need support because, as I am about to illustrate, we have a tendency to doubt whether the reasons we are responding to are ‘real.’ These reasons are, for a variety of reasons, often hard to see; and we have ample motivation to avoid seeing them. Huck Finn may have been able to do right by Jim in spite of his bad ethical theories, but how much easier would it have been for him to make the right decision with the benefit of a better theory. Our intuitions regarding the ethics of aspirational interaction are, likewise, in need of a better theory. I want, in this concluding discussion, to articulate some of these intuitions in the domain that figured so largely at the opening of this book: childbearing. I hope thereby to reveal how much it matters, even at the level of our day-to-day treatment of those we love, that we be in a position to acknowledge the reality of proleptic reasons.

In the first chapter, we considered a pair of suggestions as to how a decision theorist might explain the making of a life-changing decision, such as the choice to have children. Ullmann-Margalit considered such choices from the third person vantage point. She

determined that there was no rational way of adjudicating the debate between “Old Person,” with his current set of preferences to avoid boring family life, and “New Person,” who, having had his preferences deeply changed by the advent of children, embraces the conventionalism spurned by Old Person. Ullmann-Margalit argues that one cannot “choose” which of these points of view is better, one can only “pick,” as between type-identical boxes of cereal. Paul considered the same choice from the first person. She concluded that one ought to make the transformative choice if one wants to have new preferences, and otherwise one ought to refrain. I want to return to the question of the rationality of the childbearing decision, considered now from the point of view we have neglected: the distinctively ethical second person point of view. How should we respond or react—as we all must, at some point or another—to friends and family members traveling any distance down the road of childbearing? I have argued, above, that the aspirational approach to large scale transformative pursuits offers us better answers to both the first and the third personal deliberative questions than those offered by Paul and Ullmann-Margalit, respectively. But our discussion of self-creation and the case of Loskarn has given us reason to think that it might be in the second personal domain that the approach for which I have been arguing will really shine.

I want to begin by considering how Paul or Ullmann-Margalit might approach the second personal question. In her article, ‘What You Can’t Expect When You’re Expecting,’ Paul develops the consequences of her account of what she calls transformative choice in the area of parenthood specifically. She argues that because having a child is a transformative experience, someone who is infertile cannot know what he or she is missing. She concludes that such a person has no reason to grieve his or her infertility as a loss. She holds that it is rational to be upset over not having X only if you can anticipate the value of X for you, which is to say, if, knowing what you would get out of it, you know why you want it. For Paul, this is a matter of being able to imaginatively project for yourself the future of which you take yourself to be deprived. Since prospective parents cannot do this, it is not rational for them to be unhappy to discover that they cannot have children:

“My argument also has consequences for those who want to be able to physically conceive, carry and give birth to a child, but are unable to do so. If you want to have a child because you think having a child will maximize the values of your personal phenomenological preferences, and as a result of your inability to have a child (and thus your inability to satisfy these preferences) you experience deep sadness, depression, or other negative emotions, my argument implies that your response is not rational. This is disturbing and some might find it offensive, but it is true. Such a response is not rational. That does not mean your response is wrong, or blameworthy, or subjectively unreasonable.” (2015, p.21)

Some readers might pause over Paul’s description of the parent’s rationale for wanting children as “think[ing that] having a child will maximize the values of your personal phenomenological preferences.” Recall that for Paul, it is a general truth about decision that “To choose rationally....you choose the act that brings about the outcome with the highest estimated expected value. In the case where you have a child, the relevant

outcomes are phenomenal outcomes concerning what it is like for you to have your child..." (2015, p.5) Thus when Paul adverts to those who choose to have children in order to maximize their preferences, she intends to rule out those who may want to have children for some external reason, such as in order to have more hands to bring in the harvest. *Those* people can rationally choose to have children. But when she refers to people who have children in order to "maximize... personal phenomenological preferences," she is describing what is, in a relatively affluent society, the standard reason for wanting to have children: for its own sake.

We can set aside the question of how exactly to characterize this standard case, and whether it is right to gloss wanting children for its own sake as seeking to maximize preference satisfaction. Paul's point is that if a person who is motivated to have children in the ordinary way is unable to have children, any "negative emotions" he or she might experience at this prospect are irrational. Though Paul softens her conclusion by denying that such grief is "wrong, or blameworthy, or subjectively unreasonable," she presumably does mean to contrast this kind of case with one in which people *do* have a real reason to be upset. She is thus claiming that grieving infertile people are confused, by contrast with those who are upset at the loss of—to demonstrate the range of cases—a large sum of money, the use of a limb, or an actual, living child. The irrationality is that of being upset over a state of affairs that they cannot see as bad for them. We might compare them to a child who is upset at being denied her favorite plate to eat on. The child does not have any real reason to be upset, though we might find her response—given that it is her favorite plate—to be subjectively reasonable, and we might refrain—given that she is a child—from calling her response blameworthy or wrong.

There is a parallel line of argument on the side of those who don't want children. Paul is equally bound to charge as irrational the person who *doesn't* want children because she¹³² wants to go to college. Such a person is wrong to think that she knows what it will be like either to have a child, or to go to college; and she should not be trying to make her decision on the basis of any such projection. She cannot anticipate what that will be like to experience either the love she will feel for her child, or the intellectual excitement she will feel in college. And if, for instance, she *must* have the child, because abortion is in one way or another unavailable to her, it is, by Paul's argument, *irrational* of her to grieve the loss of her opportunity for an education.

Paul anticipates that readers will resist her conclusion, as I do myself. I think that if we think of the grief of our friends or loved ones in this way, we will treat them badly. And we will do so because we will be mistakenly blind to their predicament, which is to say, to the reasons governing their sadness. The aspiring parent does have reason—proleptic reason—to mourn her lack of access to the value in question; as the aspiring college student has reason—proleptic reason—to bemoan her lack of access to an abortion. The

¹³² From this point on I will speak, as Paul does, about motherhood specifically. Much of what I have to say will hold true to some degree for men, but the points I make here are starker when considered with reference to women. For both the burdens of childraising and the grief of childlessness characteristically fall harder on women than they do on men. It has been, and continues to be, a social and cultural fact that such issues touch women and men unequally. (See May, who notes, for instance, that "Childlessness is measured in terms of women. I found no data on childlessness among men." p.11)

fact that the person doesn't have full access to the value of what they have lost doesn't remove the sting of loss. Rather, the person's pain has a distinctive, aspirational character, in that what one mourns is never getting the chance to learn how valuable that child, or that education, would have been. One bemoans the life one never got to know.

Now let us turn to Ullmann-Margalit, who analyses such decisions as cases of 'picking.' Her account of the (ir)rational structure of what she calls 'big' decisions is also an unsatisfactory basis for a sensitive response to the grief of the unhappily pregnant, or the unhappily childless. It does not make sense to grieve the absence of a freedom to pick; if one is reduced to picking between options, then the ability to make the selection is not one that a person can reasonably prize. One is not disempowered in virtue of the fact that the supermarket shelf contains only one token of the type of cereal one prefers. The same is true of a weightier case, such as the college student discussed above (see p.37).

Suppose she would have picked (and not chosen) college A over college B. They are both good schools, and she sees that neither has more to offer her than the other. She might rationally grieve the rejection from B on the grounds that it reflects poorly on her academic worth (if she had reason to believe that), but she cannot mourn her rejection on the grounds that she would have attended B. Since she would merely have *picked* B over A, her rejection from B does not deprive her of anything. She doesn't have a real reason to be upset about where she's going to college. Things would be different if she were not able to attend any college—this is a transformative experience whose loss would be a real loss to her. And this is so even if—indeed, it is true in part *because*—she cannot quite see what she would be missing.

We want to be able to distinguish cases in which one is deprived of a picking experience, and ones in which one is deprived of an aspirational experience. Assimilating aspirants to pickers trivializes the losses experienced by the former when they are cut off from aspirational processes. It is true that both aspirants and pickers are in a situation where they do not know what they are missing. The difference between them is that aspirants were in the process of trying to learn what they were missing. The fact that the erstwhile aspirant (knows that she) will never know what she is missing is a legitimate source of grief for her.

Being cut off from one's aspiration gives one reason to lament because the aspirant's valuational condition was predicated on the value she now knows she won't acquire. Proleptic reasoners act for the sake of a future apprehension of value. They are counting on arriving at a valuational condition, exclusion from which may threaten their grip on value altogether. When such a person feels lost, empty and adrift, her feeling is a truthful reflection of what has happened to her. There is a distinctive kind of sadness appropriate to losing something you were only starting to try to get to know. This sadness is a rational response to the experience of infertility—as well as to the experience of having to abandon one's educational aspirations for motherhood. The aspiring college student who must give up those dreams to raise a child is liable to feel that she was *counting* on the college experience even as a source of meaning in her life. She doesn't know who she is if she must depart from her educational trajectory.

The historian Elaine May reports on the prevalence of a parallel sentiment among the

infertile. In the course of writing her book on infertility in postwar America, she placed an ‘author’s query’ in a variety of newspapers and magazines in which she asked childless people to write to her of their experiences. In the book, *Barren in the Promised Land*, she presents representative samples of the more than 500 letters she received.

“Marsha” (names are anonymized) writes to her,

“The grief—the loss...I spent six years of my life trying to be a mom, and it was beyond my control. For a while I couldn’t look ahead. I thought, how do I define myself if I don’t do this? What am I if not a parent?”

It does not help to tell Marsha that *she never was* a parent, and that she is exactly the person she was before embarking on her six year fertility journey. For she is not the same person, having put her other values on hold in preparing herself for motherhood. When she laments the six years she spend trying to have a child, she is not complaining about the opportunity cost. Her point is not that she wishes she would have made advances in her career or hobbies over that time. She no longer cares about the other things she could have spent that time doing, because the emotional investment into the values of motherhood has sapped her interest in anything else. This sentiment is reflected in the trajectory of another respondent. “Amanda” had initially been ambivalent about parenthood, and occupied by a demanding job as a legal secretary. Nonetheless over the course of treatment she loses interest in her job, and when the process fails she describes her condition:

“I find myself wandering [sic] every day about what I will do with my life and how I will fill this void in my soul...I want to find fulfillment in something else one day, but for now I realize that I must first heal and grieve.”

Amanda’s typological mistake—“wandering” for “wondering”—is evocative of a Freudian slip. For she describes herself as lacking anything of value, meaning or interest by which she might orient her passage through life. She has quit her job, she has given up on motherhood, and feels that there is nothing but “void” in her soul. She has the sense that there will be some valuational future for her, but at the moment she cannot anticipate it. She may well be right that in order to move forward, she must first reconcile herself to what she has lost (“I must first heal and grieve”).

If we were Amanda’s friend it would not merely be cold or ineffective to tell her that her grief is irrational. It would be false, since she does have a reason to be upset. She is mourning the loss of her own self. Amanda’s feeling of being bereft of value is of a piece with Marsha’s feeling that she lacks an identity. One’s deepest values, interests and desires—which, for these women, have all become focused onto the project of motherhood—form the core of one’s own sense of the person one is. In acknowledging her reason to be upset, we acknowledge that she has lost the person she was trying to become, and thereby the person she now is.

Aspirants open themselves up to a distinctive experience of losing everything without seeming to have lost anything at all. Because the aspirant’s value-condition is, by her own lights, derivative of the one she aspires to have, she is vulnerable to the experience

of having the valuational rug pulled out from under her. Aspirants have, to varying degrees¹³³, and in varying ways, put down roots in a possible world.

For the sake of simplicity, I have until this point assimilated the inability to have one's own biological children and the inability to adopt into a single condition of inability to have children (in any way whatsoever). Paul's argument is not restricted to counting as irrational those who grieve the lack of biological children while being able to adopt¹³⁴. The difference between the two forms of inability is immaterial to her charge of irrationality. But it is an important distinction for understanding the varieties of aspirational vulnerability, and of seeing the ways in which our own responses to aspirants may become deficient in fellow feeling.¹³⁵ I recall with shame my first conversation with a close friend who was reeling from the shock of her newly discovered infertility. I found myself driven to raise the possibility of adoption at a point where I should have recognized that topic as premature. Indeed, I believe I did recognize that the words coming out of my mouth were wrong, even cruel. What drove me to imply, as I could see I was doing, that it is self-indulgent to be attached to the idea of having one's own biological children? Why did I insinuate that it is a mark of wanting children for the 'right reasons'—whatever those are—that one is happy to come by them whichever way? The heartlessness and ignorance of such a response is heightened when it comes from a person such as myself at the time of my friend's discovery, who has had the privilege of easily conceiving and bearing their own biological children. Those willing to adopt, but wishing to adopt infants as opposed to older children, can come in for the same sort of criticism. Why do we find it so natural to task infertile people with curing overpopulation or the plight of foster children?¹³⁶

I think a certain form of callousness is engendered by the difficulty of theorizing or rationalizing proleptic reasons. One form this takes is an unwillingness to recognize aspirational loss as real—we treat aspirants as though they had *no* access to the value in

¹³³ Thus we can also use the aspirational framework to understand the grief of women whose pregnancies do not lead to motherhood. Even those women who believe that life begins at birth rather than conception experience miscarriage (especially late term miscarriage) or stillbirth as something akin to the death of a child (Aloi, 2009). Regardless of when one thinks life begins, emotional preparations for it tend to begin well in advance of that. Likewise, women who relinquish their babies for adoption often experience the event as a lifelong wound. The long-lasting grief of birthmothers is well-documented in the psychological literature, as is the fact that such grief is what Doka (1989) calls "disenfranchised": "As the child did not die, the birthmother is not expected to mourn. The perception that since the baby was put up for adoption and, therefore, a degree of comfort exists with the decision, further negates the need for grieving.... There is no public announcement of the pregnancy, the birth or the loss. The relinquishment is not even seen as a loss." (Aloi, p.28) A philosophical framework which allows us to see the grief of child relinquishment, pregnancy loss, and infertility as responsive to real—if proleptic—reasons will also expose our reason to provide these women with personal as well as institutional assistance in expressing and working through such grief.

¹³⁴ Or, for that matter, the reverse: those who are unhappy over being unable to adopt and who could but are unwilling to have biological children.

¹³⁵ May reports that the following sentiment, expressed by one of the infertile women who responded to her request for experiences, was echoed by many others: "the most difficult part of being childless and infertile was the things that friends, family and strangers said to us." (p. 13)

¹³⁶ I am setting aside the fact that such advice often wildly underestimates the difficulty and expense of adoption.

question. So, for instance, May writes that “Deirdre Kearney was one among many who provided a list of frequently heard offensive remarks: “I really would like to write a book entitled *What NOT to Say to Infertile Couples!* ‘Just forget it and go on.’ ‘I don’t think it would have bothered me if I couldn’t have children.’ ” The person who made that last comment, who I’ll call “Bill,” would presumably experience, e.g., losing custody of his children as a profound loss. Now that Bill has children, he values being their parent. But he denies that he had access to this value in advance. He thereby denies that the inability to be a parent is a loss for those who have never been parents.

The form of callousness I exhibited toward my friend took the opposite form: instead of treating her as though she had no access to reasons of motherhood, I was ignoring the proleptic character of her reason. I should have understood that my friend would long to feel her unborn baby kick, to give birth to him or her. I should have anticipated that she would be thrown by the possibility of being deprived of the experiences of pregnancy and childbirth. These desires should not be confused with those of someone (if there could be such a person) who viewed those experiences as final ends, or as intrinsically valuable. The grieving infertile person is not selfishly hyper-attuned to her own experiences. She wants to have these experiences because images of herself being pregnant and giving birth have been the (proleptic) form that her desire to be a mother has taken for her. She knows that being a mother is not the same thing as being pregnant, but for her, for now, her limited grasp of the value does not allow her to separate those two conditions. Being pregnant or nursing a newborn may be as far ahead as she can see into the motherhood project. It is by looking forward to these experiences that she has an antecedent grasp on the good of parenthood. (I’d wager that few prospective parents fantasize about parenting teenagers.)

Bill’s response to Deirdre and my response to my friend share a blindness to the way in which the infertile person is antecedently invested in motherhood. If having a child were like riding a rollercoaster or tasting a new flavor—something you did for the sake of having a new experience—Paul would be right that deprivation makes no mark on the deprived subject. Someone cannot really be ‘robbed’ of such an experience, because he cannot feel the loss. If you have never tasted chocolate, then a chocolate lover such as myself might jokingly say ‘I feel sorry for you.’ But it is I, and not you, who have reason to be distraught over the prospect of the end of the world’s chocolate supply. The same holds for profound life changes one had no interest in: someone who was fully and strictly indifferent to the project of becoming a mother, appreciating opera, traveling to Egypt, or becoming an astronaut, would have no reason to be upset if the path toward any of these pursuits were foreclosed to her¹³⁷.

Alternatively, suppose having a child were a mid-sized decision such as buying a car. S is looking to buy a car in order to get to and from work more easily. It would be perfectly intelligible to raise the possibility of his moving closer to work in order to be in

¹³⁷ Though it is a well recognized phenomenon that people will discover the limits of their indifference in such a scenario, coming to long for the very option that is now foreclosed. Sometimes this represents a ‘grass is greener’ form of irrationality; at other times, I suspect, one simply learns that one really did want something one took oneself not to want.

a position to rely on public transportation. It won't usually be¹³⁸ out of place to point out incidental benefits of this alternative option, such as mitigating environmental damage, living closer to the action of the city, getting more exercise on the way to the office, etc. S might be happy to combine a solution to his transportation problem with other problems, such as: the social life problem, the environmental problem, the healthy lifestyle problem. If having a child were like buying a new car, it would be appropriate to point out to someone who has just learned that she will have difficulty having children without extreme, expensive medical intervention that she can contribute to solving the foster child problem; that having a newborn will interfere with sleep, and therefore with her ability to succeed at her job; that pregnancy is awful for one's teeth and for one's general health, etc.

But this deliberative presentation of the varieties of parenthood will typically be the wrong way to talk to someone first facing the prospect of infertility. The problem is not that the infertile person is being asked to weigh something important—her future child—against things too trivial to matter in the face of that value. Her health, her job, her finances and the plight of unwanted children might be deeply important to her, both pre- and post- parenthood. Most parents will at some point have to carefully weigh some dimension of their child's well being against the importance of helping the less fortunate, or their own health and career. The problem is not that the value of the nonexistent child somehow dwarfs all value. The problem is that the prospective parent doesn't yet know how to weigh all these things against one another. Her grasp on the value of parenthood is not like S's grasp on the value of transportation. She does, of course, currently care about the plight of the less fortunate, her own health, job, finances, etc. But her concern for those things has developed over her years as a nonparent. She knows how someone who doesn't have children values her job, her health, her finances, and her moral obligations to the less fortunate. She doesn't have available to her the new way of thinking about these values when they are combined with the valuation of one's child. And she won't, not until she actually is a parent. And that's a merely necessary condition: if she is like me, she may not feel at home in motherhood until some time after she is taken by the world at large to be one. The transition into motherhood does not involve losing hold of all the things one used to value, but it may shake one's antecedent grip on those values. I became unsure *how* to value the things I used to value—how to value money, and health, and morality “as a mother.” Reconstructing one's values takes time; large scale transformative pursuits often involve a kind of rebirth even with respect to those values that straddle the transformative event.

When we point out to the person who has just learned she is infertile the external selling points of adoption we are bulldozing over the multiple evaluative perspectives that divide her deliberative faculty. We are treating her as though she had a unity she doesn't yet have; and the reason she doesn't have it is precisely that she is trying so hard to become a parent. We are ignoring how hard she has to work to even see the value—that of

¹³⁸ Though it also might be. You might know full well that he was looking forward to buying this car for years as his first ever big, independent purchase; it might have important psychological place in his move toward adulthood. It might still be right to advise him not to buy it, but one might need to put the point with more sensitivity than the conversation described above suggests.

parenthood—that we are taking her to have reflexive access to. We are ignoring how vulnerable her valuational position, which is to say, her own sense of her identity, currently is.

A proleptic reason poses a distinctive expository difficulty that we exploit in our insensitive responses to aspirants. Reasons are answers to “why” questions—either “why do you intend to ϕ ?” or “why did you ϕ ?” An aspirant is not in an ideal position to answer such a question—for her reason always has a proximate face, citation of which represents her activity as lesser than it is, and a distal face, citation of which ennoble her current activity beyond its rightful status. The music student cannot say that she’s taking the class for the grade, nor can she say she’s taking it because music is intrinsically valuable. The infertile person can’t say she wants, for purely selfish reasons, to have the phenomenal experience of pregnancy; nor can she say, as she will be able to, that she’s acting out of parental love. A proleptic reasoner will have trouble explaining exactly why she is doing what she is doing, though once she gets to her destination she will say, “this was why.” Her current rational understanding of herself is predicated on the better understanding she looks forward to having one day. If we want the best justification for her actions, we have to turn to her future self.

All of this is, at least in the case of parenthood, quite salutary. It would be terrible if we wanted children for (what we could antecedently appreciate as) good reasons. For what we look forward to in looking forward to becoming parents is the activity of loving some specific child for his or her own sake. But, whether adopting or not, no one can know in advance which person she will get. We must be open to loving any child, but that does not mean that we must love every child. We aren’t required to harbor a love for babies or children in general, as might be a job requirement for a nursery school teacher. We must be open to loving some one (or a few) specific children—exactly those children and no others. And so the content of the attitude by which we move ourselves toward parenthood must be indeterminate, but determinable. It is good that we enter parenthood for the most inchoate of reasons; for that puts us in a position to let our children fill out what parenthood means for us.

END

Bibliography

Aloi, J. A., 2009: "Nursing the disenfranchised: women who have relinquished an infant for adoption *Journal of Psychiatric and Mental Health Nursing*, 2009, **16**, 27–31

Anscombe, G. E. M. 1963. *Intention*. Second edition. Oxford: Basil Blackwell.

Aristotle 1980: *Ethica Nicomachea*. (Ingram Bywater, ed., Oxford: Oxford University Press.)

Arpaly, Nomy. 2003: *Unprincipled Virtue: An Inquiry Into Moral Agency*. (New York: Oxford University Press.)

Anderson, E. 1997: "Practical Reason and Incommensurable Goods", in *Incommensurability, Incomparability and Practical Reason*, (ed. R. Chang) Cambridge, Mass: Harvard University Press.

Augustine, 1993: *On Free Choice of the Will*, trans. Thomas Williams (Hackett, IN).

Berofsky, B. 1995: *Liberation from self: A Theory of Personal Autonomy*, Cambridge University Press.

Barnes, J. ed., 1984: *The Complete Works of Aristotle: The Revised Oxford Translation* (Princeton: Princeton University Press).

Bratman, Michael. 1979: "Practical Reasoning and Weakness of the Will," *Nous* 13.
_____ 1987: *Intention, Plans and Practical Reason*. (Chicago: University of Chicago Press.)

Bratman, M. 2003: "Autonomy and Hierarchy," *Social Philosophy and Policy*.

Broadie, S. 1991: *Ethics with Aristotle*. New York.

Broadie, S., ed. 2002: *Aristotle, Nicomachean Ethics* trans. Christopher Rowe, (Oxford: Oxford University Press.)

Burnyeat, M. Aristotle On Learning to Be Good. In Amélie Oksenberg Rorty, ed., *Essays On Aristotle's Ethics*, University of California Press: Berkeley, 1980, pp. 69-92.

Buss, S. 2012: "Self Determination in the Passive Mode," *Ethics*.

Buss, S. ms. "Why a Determinate Constitutive Aim is Incompatible with the Conditions of Rational Agency."

Callard, A. 2014: Ignorance and Akrasia Denial in the *Protagoras*," *Oxford Studies in Ancient Philosophy*.

Callard, A. 2015: Review of Adrienne Martin, *How We Hope: A Moral Psychology*, *Ethics*.

Callard, A. : 2015 “The Weaker Reason,” *Harvard Review of Philosophy* vol. 22.

Callard, A., “Liberal Education and the Possibility of Valuational Progress,” under review.

Callard, A. *Proleptic Reasons* forthcoming in *Oxford Studies in Meta-Ethics* vol. 11.

Chang, R. 1997: ‘Introduction,’ in *Incommensurability, Incomparability and Practical Reason*, (ed. R. Chang) Cambridge, Mass: Harvard University Press

Clarke, Randolph. 2003: *Libertarian Accounts of Free Will* (Oxford: Oxford University Press).

Clarke, Randolph 2005: “On an Argument for the Impossibility of Moral Responsibility,” *Midwest Studies in Philosophy*, 29 pp. 13– 24.

Coope, U. 2012: “Why does Aristotle think that Ethical Virtue is Required for Practical Wisdom?” *Phronesis* 57 pp. 142-163.

Cooper, J. ed., 1997: *Plato, Complete Works*. (Hackett : Indianapolis)

Davidson, D. 1980¹: “Actions, Reasons and Causes,” in Donald Davidson, *Actions and Events*. (Oxford: Clarendon Press.)

Davidson, D. 1980²: “How is Weakness of the Will Possible?” in Donald Davidson, *Actions and Events*. (Oxford: Clarendon Press.)

Davidson, D. 1985: “Deception and Division,” in E. Lepore and B. McLaughlin (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. (New York: Blackwell.)

Dickens, C. *A Tale of Two Cities*

Doka K.J., ed. (1989) *Disenfranchised Grief: Recognizing Hidden Sorrow*. Lexington Books, New York.

Doris, J. 2002: *Lack of Character*, Cambridge University Press.

Dworkin, G. 1976: “Autonomy and behavior control,” *The Hastings Center Report*.

Ekenberg, T. 2015: “Voluntary Action and Rational Sin in Anselm of Canterbury,” *the British Journal for the History of Philosophy*.

- Ekstrom**, L. 1993: "A Coherence Theory of Autonomy," *Philosophy and Phenomenological Research*.
- Elster**, J. 1982: "Sour Grapes: Utilitarianism and the Genesis of Wants," in Amartya Sen & Bernard Williams (eds.), *Utilitarianism and Beyond*. Cambridge University Press.
- Fischer**, J. 2006: "The Cards that Are Dealt You," *Ethics*.
- Frankfurt**, H. 1971: "Freedom of the Will and the Concept of a Person," *Journal of Philosophy*.
- Frankfurt**, H. 1978: "The Problem of Action", *American Philosophical Quarterly*, 15(2): 157–62.
- Frankfurt**, H. 1976: "Identification and Externality" in Amelie Rorty, ed., *The Identities of Persons* Berkeley University of California Press 1976.
- Frankfurt**, H. 1988: "Identification and Wholeheartedness," in his *The Importance of What We Care About* (Cambridge University Press).
- Frede**, M. and Long, A., 2012: *A Free Will: Origins of the Notion in Ancient Thought* Sather Classical Lectures, 68 (Berkeley: University of California Press).
- Garnett, M. 2013: "The Autonomous Life: A Pure Social View," *Australasian Journal of Philosophy*.
- Goodfellas*, 1990: Directed by Martin Scorsese, Warner Brothers Films.
- Hardy**, G. 2005 (1940): *A Mathematician's Apology*, published by the University of Alberta Mathematical Sciences Society, available since 2005 at <http://www.math.ualberta.ca/mss/>
- Harman**, E. 2015 "Transformative Experiences and Reliance on Moral Testimony," *Res Philosophica*
- Harman**, G. 2003: "No Character or Personality," *Business Ethics Quarterly*.
- Harman**, G. 2000: "Desired Desires," in his *Explaining Value and Other Essays in Moral Philosophy* (Oxford: Clarendon Press, 2000), pp. 117–136, at p. 135.
- Holton**, R. 2004: "Rational Resolve" (RR) *Philosophical Review* 113 (4).
- Irwin**, T. H. 1999: *Aristotle: Nicomachean Ethics*. Translated, with introduction, notes, and glossary.(Hackett: Indianapolis).
- Jackson**, F. 1984: "Weakness of Will," *Mind* vol. 93, no. 369.

- Jimenez, M.** 2016: "Aristotle on Becoming Virtuous by Doing Virtuous Actions," *Phronesis*.
- Jones, Karen.** 2003: "Emotion, Weakness of Will and the Normative Conception of agency" in A. Hatzimoysis (ed.) *Philosophy and the Emotions*. (Cambridge: Cambridge University Press.)
- Kamtekar, R.** 2004: "Situationism and Virtue Ethics on the Content of Our Character," *Ethics*.
- King, S.** 2000: *On Writing*, (NYC: Simon and Schuster)
- Kolodny, Niko** "Love as Valuing a Relationship," *Philosophical Review* 112 (2):135-189 (2003)
- Kolodny, Niko.** 2005: "Why Be Rational?" *Mind* vol 114, no. 455.
- Korsgaard, C.** 1986: "Skepticism about Practical Reason," *The Journal of Philosophy*, Vol. 83, No. 1., pp. 5-25.
- Korsgaard, C.** 1996: *The Sources of Normativity*, Cambridge University Press.
- Lawler, J.** 2013: *For Jessica* Lawler and Daughter Publishing
- Leiter, B.** 2011: "The Paradox of Fatalism and Self-Creation in Nietzsche," in John Richardson and Brian Leiter, eds., *Nietzsche*, Oxford University Press.
- Lewis, David.** 1989: "Dispositional Theories of Value." *The Proceedings of the Aristotelian Society, Supplementary*. Volume 63: 113-137.
- Little, M.** "Abortion and the Margins of Personhood," *Rutgers Law Journal* vol. 39 2008
- Marcus, R.** 2014: "The Tragedy of Ryan Loskarn," *The Washington Post*, Feb. 16th.
- Markovits, J.** 2010. "Acting for the Right Reasons," *Philosophical Review*, 119(2), 201–42.
- Markovits, J.** 2011a. "Why Be an Internalist about Reasons?" in *Oxford Studies in Metaethics: Volume 6*. Oxford: Oxford University Press, 255–79.
- Markovits, J.** 2011b. "Internal Reasons and the Motivating Intuition." In Michael Brady (ed.), *New Waves in Metaethics*. London: Palgrave Macmillan, 141–65.
- Martin, A.** 2013: *How We Hope: A Moral Psychology* (Princeton: Princeton University Press)

- May, E.** 1995: *Barren in the Promised Land* (Basic Books, New York, NY).
- McDowell, J.** 1998: "Virtue and Reason," in his *Mind, Value and Reality* (Cambridge, Mass.: Harvard University Press)
- McDowell,** 1998: "Some Issues in Aristotle's Moral Psychology," in his *Mind Value and Reality* Harvard University Press.
- McGrath, S.** 2011: "Skepticism about Moral Expertise as a Puzzle for Moral Realism," *Journal of Philosophy* 108: 111-137.
- McIntyre, Alison.** 1993: "Is Akratic Action Always Irrational?" in O . Flanagan and A. Rorty (eds.), *Identity, Character and Morality* (Cambridge, Mass.: MIT Press.)
- Mele, A.** (1995). *Autonomous agents*. New York: Oxford University Press.
- Millgram, E.** 1997: *Practical Induction*, Harvard University Press.
- Moran, R.** 2002: 'Frankfurt on Identification' in *Contours of Agency: Essays on Themes from Harry Frankfurt*, Buss, Sarah (ed.), Bradford Book, MIT Press: Cambridge MA: 189-217.
- Murdoch, I.** 1996: "Metaphysics and Ethics" Schweiker (eds.), *Iris Murdoch and the Search for Human Goodness* (Chicago: University of Chicago Press), p.252.
in Maria Antonaccio and William
- Nielsen, K.** 2011: "Deliberation as Inquiry: Aristotle's Alternative to the Presumption of Open Alternatives" *The Philosophical Review* vol. 120 .no. 3
- Nietzsche, F. 1966: *Beyond Good and Evil*, trans. Walter Kaufmann, Random House.
- Noggle, R.** 2005: "Autonomy and the Paradox of Self-Creation" in James Taylor, ed., *Personal Autonomy*, Cambridge University Press.
- Owen, D. and Ridley, A.** 2003: "On Fate," *International Studies in Philosophy*.
- Parfit, D.** 2011: *On What Matters* vol. 1 (Oxford: Oxford University Press)
- Paul, L. A.** 2014: *Transformative Experience*. Oxford: Oxford University Press.
- Paul, L.** 2015: "What you can't expect when you're expecting," *Res Philosophica* 92:2.
- Paul, S.** 2011 Deviant Formal Causation JOURNAL OF ETHICS & SOCIAL PHILOSOPHY

- Pettigrew, R.** Transformative experience and decision theory," forthcoming in *Philosophy and Phenomenological Research*.
- Rahman, Z.** 2014: *In the Light of What We Know* (Farrar, Strauss and Giroux: New York, NY)
- Raz, J.** 1988: *The Morality of Freedom*, Oxford University Press.
- Railton, P.** 1992: "Pluralism, Determinacy, and Dilemma," *Ethics* 102.
- Rosati, C.** 1995: "Persons, Perspectives and Full Information Accounts of the Good," *Ethics* 105.
- Scheffler, S.** 2010 "Valuing," in his *Equality and Tradition*, pp. 15-40 (Oxford: OUP)
- Schillebeeckx, E.** 1969: *God and Man*, trans. Edward Fitzgerald and Peter Tomlinson, (New York: Sheed and Ward).
- Schmidtz, D.** 1994: "Choosing Ends," *Ethics* (Vol. 104, No. 2), pp. 226-251
- Schroeder T.** and Arpaly N. 1999: "Alienation and Externality" *Canadian Journal of Philosophy*.
- Setiya, K.** 2007: *Reasons without Rationalism*, Princeton University Press.
- Smith, A.** 2005: "Responsibility for Attitudes: Activity and Passivity in Mental Life," *Ethics*.
- Smith, M.** 1995: "Internal Reasons," *Philosophy and Phenomenological Research* vol. LV no. 1, pp. 109-131.
- Smith, M.** 1987: "The Humean Theory of Motivation," *Mind* vol 96, no. 381.
- Smith, M.** 1992: "Valuing: Desiring or Believing?" in D. Charles and K. Lennon (eds)., *Reduction, Explanation, and Realism* (Oxford: Oxford University Press),
- Smith, W.** 1872: *A Smaller History of Greece, from the Earliest Times to the Roman Conquest* (Harper & Bros., New York)
- Sobel, D.** 1994: "Full Information Accounts of Well-Being," *Ethics* 104 pp.784-810
- Stocker, Michael.** 1979: "Desiring the Bad: An Essay in Moral Psychology," *The Journal of Philosophy* vol. 76, no. 12.
- _____ 1992: *Plural and Conflicting Values*. (New York: Oxford University Press.)

Strawson, G. The Impossibility of Moral Responsibility Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, Vol. 75, 1994), pp. 5-24

Strawson, G. Freedom and Belief (Oxford: Oxford University Press, 1986).

Ullmann-Margalit, Edna and Morgenbesser, S. 1977: "Picking and Choosing," *Social Research*, Vol. 44, No. 4, pp. 757 -785.

Ullmann-Margalit, Edna. 2006. "Big Decisions: Opting, Converting, Drifting." In *Political Philosophy*, edited by Anthony O'Hear, Vol. 58 of *Royal Institute of Philosophy Supplements*, 157–172.

Valdman, M. 2011: "Autonomy History and the Origins of our Desires," *Journal of Moral Philosophy*.

Velleman, D. 2002: "Motivation by Ideal" pp.90-104 *Philosophical Explorations* 5 (2)

Velleman, D. 1998: 'Brandt's definition of good," *Phil Review* 97

Walker, Arthur. 1989: "The Problem of Weakness of Will," *Nous* vol. 23, no. 5.

Wallace, J. **The View from Here. On Affirmation, Attachment, and the Limits of Regret** (New York: Oxford University Press, 2013).

Watson, G. 1975: "Free Agency," *The Journal of Philosophy*.

_____ 1987: "Responsibility and the Limits of Evil: Variations on a Strawsonian theme" in John Martin Fischer, ed., *Perspectives on Moral Responsibility*.

Wiland, E. 2014: "Rossian Deontology and the Possibility of Moral Expertise," in *Oxford Studies in Normative Ethics*, ed. M. Timmons, Oxford University Press vol. 4

Wiggins, David. 1980:"Incommensurability: Four Proposals" in Ruth Chang, ed. *Incommensurability, Incomparability and Practical Reason*. (Cambridge, MA: Harvard University Press.)

Williams, B. 1981: "Internal and External Reasons", in his *Moral Luck* (Cambridge University Press).

Williams, B. 1981: "Persons, Character and Morality," in his *Moral Luck* (Cambridge University Press).

Williams, B. 1995, "Internal Reasons and the Obscurity of Blame," in his *Making sense of Humanity* (Cambridge: Cambridge University Press) 39-45.

Wolf, S. 1987: "Sanity and the Metaphysics of Responsibility," in Ferdinand Schoeman, ed., *Responsibility, Character, and the Emotions*, Cambridge University Press.

Wolf, S.: "Character and Responsibility," forthcoming in *The Journal of Philosophy*.

Wu, W. 2011: "Attention as Selection for Action," in *Attention: Philosophical and Psychological Essays*, Christopher Mole, Declan Smithies, and Wayne Wu, eds., (New York: Oxford University Press).