Title
Intention and Normative Belief

Author
Chislenko, Eugene

Publication Date
2016

Peer reviewed|Thesis/dissertation

Intention and Normative Belief

By

Eugene Chislenko

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Philosophy

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge:

Professor Hannah Ginsborg, Co-Chair
Professor R. Jay Wallace, Co-Chair
Professor Hubert Dreyfus
Professor Tania Lombrozo

Spring 2016

Abstract

Intention and Normative Belief

by

Eugene Chislenko

Doctor of Philosophy in Philosophy

University of California, Berkeley

Professor Hannah Ginsborg and Professor R. Jay Wallace, Co-Chairs

People can be malicious, perverse, compulsive, self-destructive, indifferent, or in conflict with their own better judgment. This much is obvious—but on many traditional views, it seems puzzling or even impossible. Many philosophers, from Plato and Aristotle to Kant, Davidson, and others, have thought that we act only "under the guise of the good," doing only what we see as good, or best, or what we ought to do. These "guise-of-the-good" views offered a way to make sense of the attribution and explanation of action, while maintaining a generous view of human nature as essentially pursuing the good. But are they not hopelessly narrow and naïve? It seems clear that we often do what we do *not* see as good, and even what we see as bad. The classical view seems to paint an impoverished picture of human life, leaving out widespread and important forms of activity. It now seems natural to give up on such a view, and to look for a more viable alternative.

In this dissertation I argue that, far from being narrow and naïve, even an ambitious "guise-of-the-good" view can offer a compelling picture of action in all its variety. In Chapter 1 I introduce the appeal, variety, and difficulties of such views, and suggest a strategy for investigating them: start with a simple, ambitious "guise-of-the-good" view, and see why and in what ways it must be weakened. The ambitious view I begin with is the view that intentions are themselves normative beliefs—that my intention to go to the store *is* a belief that I ought to go. I call this view the Identity View. In the rest of the dissertation, I argue that it does not need to be weakened at all. The Identity View can address a range of apparently powerful counterexamples and theoretical challenges, while offering a compelling conception of the nature and the details of our intentions.

A central kind of counterexample is action and intention in which we intend to do what we believe we should *not* do: eat dessert, for example, or insult a friend. Such 'akratic' action seems widespread, and difficult to account for on a guise-of-the-good view. The Identity View has an added consequence: if intentions are themselves normative beliefs, it seems we will also have *beliefs* we believe we should not have. Akratic beliefs can seem especially difficult to describe convincingly, partly since, according to many people, they are not even possible. I

consider these beliefs in Chapters 2 and 3. The first of these chapters argues that the leading arguments against the possibility of akratic belief assume what they purport to show, and are derivative expressions of an underlying puzzlement about how belief can be akratic. The second addresses the puzzlement directly, by offering a conception of akratic belief, together with a range of examples of it. I argue that the marks by which we normally attribute belief—marks such as responsiveness to evidence, felt conviction, recall in relevant circumstances, reporting to others, and use in further reasoning—can be recognized, with some complications, in akratic cases as well.

Chapter 4 turns to akratic action and intention. Here, it is widely accepted that akratic cases are possible, but this possibility itself seems inconsistent with views like the Identity View, on which we can only intend what we do believe we ought to do. But the implication of the Identity View is not that we cannot intend to do what we believe we ought not do. Instead, the implication is that, when we intend to do what we believe we ought not do, we must *also* believe that we ought to do it. I argue that akratic actions can be understood as cases of conflicting normative beliefs, in which we act on one belief while still holding the other, often more strongly or reflectively. These cases are, at the same time, cases of conflicting intentions. In the second half of the chapter, I defend the attribution of this much conflict in akratic cases.

An account of akratic belief and action is only the first step toward a defense of a guise-of-the-good view. The full range of troubling cases is much broader. We sometimes seem to act intentionally without having any belief about whether we ought to take the alternative we intend to take. At other times, it seems, we can believe we ought to do something—get out of bed, or donate to charity—and have no intention to do it. In these cases, either the intention or the corresponding belief seems entirely absent, rather than in conflict. In Chapters 5 and 6, I offer a response to these other counterexamples. The first, an apparent lack of normative belief, arises in cases of indifference, like that of Buridan's Ass, and in cases of apparently incommensurable or incomparable values. I argue in Chapter 5 that, even when we see no reason to favor one particular alternative over others, we can decide to act nonintentionally—to "just pick" an alternative and take it. We can believe we ought to act nonintentionally, and such a belief can account for our intentions in such cases. Drawing on empirical studies in psychology, I argue in Chapter 6 that the second kind of case, known as *accidie*, is best understood through an analogy to fatigue, in which an intention is in fact present but hindered by a psychological obstacle. Typical cases of this kind are ones in which we fail to act on an intention, rather than failing to intend to do something.

Although it can be stated in one short sentence, the Identity View is the heart of a general theory of intention. It offers a way of understanding what intention is: an intention is a particular kind of belief. Chapter 7 begins to develop that theory, considering a set of more general issues about the nature of intention and belief. By addressing a series of apparent disanalogies between intention and belief with respect to 'direction of fit', voluntary control, and reasoning, I try to spell out what thinking of intention as normative belief entails, and why the view is believable. Together, these chapters offer a systematic defense of the classical thought that human motivation has an essential evaluative element. Even when confused, conflicted, or exhausted, we intend to act as we believe we should.

Preface

Moral philosophy is an escape from dogmatism. It does not just insist that we be kind, eat healthy, and save lives if we can. It asks why we should do any of the things that have been thought to be good or right. To do moral philosophy is, partly, to ask whether and how it can answer the questions it is meant to address.

Reflecting about moral questions usually turns at some point to reflecting about human nature. Plato's consideration of justice brought him to think about the nature and parts of the soul. Aristotle's inquiry into the characteristics of a good life led him to think about the characteristic activity of human beings. Hume and Kant took on a systematic investigation of willing, reasoning, and desire as part of their moral theory. Moral philosophy incorporates moral psychology, and helps make it interesting.

These two kinds of reflection are related. One way to ask whether and how moral questions can be answered is to look for a source of value, goodness, or obligation in the nature of thought or action. The hope is that an adequate moral psychology can lead to an adequate moral philosophy, by helping to show which answers to moral questions are legitimate, and why. We can call an attempt to reach a conclusion about how one should act from a view about the nature of action or thought a "foundational argument" in moral philosophy.

Foundational arguments are not new. Kant argued that "A free will and a will under moral laws are one and the same"(1997, 4:447)—that to govern oneself is to be governed by morality. As Korsgaard (2009, 32) puts it: "The laws of logic govern our thoughts because if we don't follow them we just aren't thinking…The laws of practical reason govern our actions because if we don't follow them we just aren't acting, and acting is something that we must do." For Kant and Korsgaard, particular laws, or principles, have an inescapable claim on us. If we ask why we must follow them, there is a compelling answer. We must follow these principles, because that is what it is to act.

There is widespread skepticism about the prospects of foundational arguments. The notion of action, or of thought, can seem not clear enough, or not restrictive enough, for action or thought to have any interesting basic features. And even if they do, it can seem obscure how these features could give rise to anything like an obligation to perform one action rather than another. So rather than pursuing these arguments, we could instead start with the moral judgments we already make, examine them to see whether they are consistent with each other, and try to give up some of them in order to leave the remaining ones, as much as possible, in harmony with each other. Many have thought that the role of moral theory is to help establish consistency in our moral judgments. With respect to moral foundations, this is, to a large extent, the spirit of our age: a conviction that foundational arguments are hopeless, and an interest in what can be done with more piecemeal methods.

I think this is exactly backwards: these other methods are hopeless, and we have yet to see what can be done with foundational arguments. A fully consistent set of moral intuitions is just that: consistent. Beyond consistency, these intuitions have no claim to be correct or legitimate. Nor do these intuitions do much to explain why we should be bound by consistency. And without any explanation of why one set of intuitions, consistent or not, should be followed, we are not much farther than when we started. Two people or two cultures with competing sets of intuitions can be left with little to say to each other. And the problem arises even for a single person. I can ask whether my own set of intuitions is any more than a set of intuitions, in the sense of feelings or inclinations, which I may trade in for any other if and when I can. The consistency of intuitions with each other can still leave the whole structure seeming groundless. Hearing that nothing more than consistency is attainable does not remove the sense that one is no better off than a skeptic in respectable clothing. I think that settling for consistency in moral judgment is just that: skepticism, plus a dogmatic requirement of consistency. It is an unstable strategy, with nothing to offer beyond the harmonization of the intuitions of a particular person or culture. When the culture changes, an equilibrium of intuitions becomes a relic of a bygone age.

Some intuitions, like the intuition that it is good to save a life, might be universally shared, or at least close to universal in humans as a biological species. A catalogue of these might never become just a relic. But even a set of intuitions that is universally shared does not distinguish between the intuitions that are and are not worth sharing. The history of a species leaves deeply ingrained intuitions, some of which can be important to resist. Actual intuition or adherence is a questionable foundation for a moral principle, in a way that being a necessary condition for thought or action may not be. These issues quickly become complex. But that is all the more reason to consider them in detail.

So far, there is no conclusive proof that foundational arguments can ever succeed, and no conclusive proof that they cannot. Moral theory is still in its early stages, with a great deal left to discover. But the question of the viability of a foundational argument is still open. A sense of the futility of harmonizing intuitions leads naturally to looking for a more compelling alternative.

The pursuit of a foundational moral argument has two main parts. First, it needs a conception of thought or action, from which values, obligation, or answers to practical questions might arise. Second, it needs to show how these can arise from a purely descriptive conception. These are, to put it another way, the foundations, and the transition from foundations to practical implications. In this dissertation, I take on the first part of this project. I engage in moral psychology in the service of moral philosophy. Even with ultimate failure as a foundational argument, such a project would be independently interesting, since it promises to show us something fundamental about ourselves. The hope is that its conclusions will also be morally significant. But I leave the moral implications for another time. My focus here is not on a conclusion about what we should do, but on the inescapability of making such conclusions. The question is whether beliefs about what one should do are essential to action in general.

Table of Contents

Outline of the Dissertation

Chapter 1: The Guise of the Good

Many philosophers have held that one can pursue or desire only what one sees as good—or, "under the guise of the good." This dissertation defends a version of this view.

I. Motivations

There is a variety of deep motivations both for holding a guise-of-the-good view, and for being interested in them. Some central motivations concern the nature of action, unified explanation of action and evaluation, parallels with belief, historical precedent, implications for moral theory, and generosity of interpretation.

II. Challenges

It can seem obvious that we do and want what we do not see as good. We seem to act against our own better judgment, or without one, as well as evaluate without acting. There are also more abstract challenges concerning voluntary control, direction of fit, and types of reasoning.

III. Parameters

There are countless possible guise-of-the-good views. For example: An (action / intention / desire) (is / requires / requires a capacity to have) a (belief / judgment / appearance) that the (action / end / outcome) is (what one ought to do or bring about / good / in some way good). Within each of these parameters, there can be further specifications, additions, and interactions with views on related topics.

IV. Strategies

Which variants are worth considering? Strategies for selecting a starting point include simplicity, historical precedent, connection to the motivations in §II, and ambition. One view with all of these is that intention is a belief that one ought. We can call this the Identity View.

V. Summary of the Dissertation

This dissertation defends the Identity View. Chapters 2-6 offer an account of

particularly difficult cases: akratic belief (Chapters 2-3), akratic action (Chapter 4), lack of evaluation (Chapter 5), and lack of motivation (Chapter 6). Chapter 7 addresses the abstract challenges.

Chapter 2: Believing Against One's Better Judgment, I: Impossibility Arguments

I. The Problem

If we can have intentions we believe we ought not have, then on the Identity View, we can have beliefs we believe we ought not have. But such 'akratic' beliefs are often thought to be impossible.

II. The Nullification Argument

According to one argument, belief in conclusive reason against some belief 'nullifies' the force of any apparent reason to have the belief. This argument applies to only some akratic beliefs, faces powerful counterexamples, and offers no independent reason to think that beliefs cannot be akratic.

III. The Argument from Moore's Paradox

Akratic belief can seem as absurd as believing: "It's raining, but I don't believe it." But even if such 'Moorean' belief is impossible, akratic belief does not require conjunctive beliefs like: "It's raining, but I shouldn't believe it." And even these conjunctive beliefs are importantly different from Moorean ones.

IV. The Argument from Transparency

Beliefs are normally 'transparent', in the sense that each of us comes to a belief about whether she believes that *p* by coming to a belief about whether *p*. But this transparency can fail, and does not offer a compelling argument against the possibility of beliefs one believes one should not have.

Chapter 3: Believing Against One's Better Judgment, II: How Akratic Belief is Possible

The impossibility arguments seem powerful because they express an underlying puzzlement. This chapter offers a conception of akratic belief to address that puzzlement.

I. Rorty's Catalogue

Amélie Rorty describes a "catalogue" of akratic beliefs, including "intellectual," "interpretative," "inferential," and "predictive" *akrasia*. Her method of "distinguishing the strands" in belief is useful, but she assumes that akratic beliefs must be voluntary.

II. Scanlon's Dispositional View

For T.M. Scanlon, "Belief is not just a matter of judgment but of the connections, over time, between this judgment and dispositions to feel conviction, to recall as relevant, to employ as a premise in further reasoning, and so on…. Akrasia involves the failure of these connections."

III. How is Akratic Belief Possible?

Combining Rorty's and Scanlon's views, we can distinguish several marks of belief: sensitivity to evidence, recall, felt conviction, reporting to others, and use in further reasoning. Using these marks, we can sometimes recognize both on particular belief, and a second belief that we should not have the first one.

IV. Why is Akratic Belief Puzzling?

It is puzzling, partly in the ways akratic action is puzzling; partly because it is a state; partly because it might, for contingent reasons, be relatively rare; partly due to our limited imagination; and partly because of our prior theoretical commitments.

Chapter 4: Acting Against One's Better Judgment

It seems clear that we often intend to do things we believe we ought not do, and often act on such intentions. But on the Identity View, such 'akratic' intentions and actions seem impossible.

I. The Possibility of Conflicting Beliefs

The Identity View does not have the implication that akratic action and intention are impossible. We can intend and do what we believe we ought not do, if we *also* believe that we ought. We should allow that such conflicts are possible.

II. The Conflicting Belief View

What the Identity View entails is the Conflicting Belief View: all akratic action and

intention involve conflicting beliefs about what one ought to do. As with akratic belief, an understanding of conflict in belief can explain akratic action and intention.

III. Objections and Replies

Appealing to conflicting beliefs reduces one problem to another; helps explain why akratic action is puzzling; attributes a small but plausible amount of error in self-attribution; and leaves room for various forms of asymmetry in belief.

V. The 'Better' Intention

Akratic intention can be understood as a conflict not only between beliefs, but also between intentions. Seeing this helps us see how the conflict is 'practical'.

Chapter 5: Motivation without Evaluation

In other cases, we can believe our alternatives are equally good, or be unable to compare them. It seems we must then intend and do something without any belief at all about whether we ought to.

I. Buridan's Ass

Buridan's Ass starves to death, unable to choose between identical bales of hay. We face and successfully resolve many such cases, apparently without believing we ought to take the alternative we take.

II. Initial Responses

One can deny the possibility or resolution of Buridan cases, or insist that we choose at random or just let our attention fall on one alternative. All of these responses leave at least some Buridan cases unexplained. But they show what an adequate response must do: e.g., give a recognizable account that leaves no clear counterexamples.

III. Deciding to Act Non-Intentionally

In resolving Buridan cases, we can simply decide to act non-intentionally. This kind of resolution is recognizable and always available. And once we do it, we do have an alternative we can believe we ought to take. This keeps us from ever being forced to form an intention without such a belief.

Chapter 7: Intention as Normative Belief

I. Toward a General Theory

The Identity View is the heart of a general theory of intention. This chapter begins to develop that theory, partly by addressing some challenges to it.

II. Direction of Fit

Beliefs aim to match the world; intentions aim to make the world match them. If they have different 'directions' of matching or fit, they cannot be identical. But a belief that one ought to do something, and an intention to do it, both aim to fit whether one ought to, and both aim to make one's actions fit them.

III. Voluntary Control

Beliefs seem distinctive in that we cannot believe at will; we cannot simply choose what to believe. On the Identity View, intentions are beliefs, so this contrast between belief and choice seems lost. But in the relevant sense, we cannot intend at will, either.

IV. Reasoning

The Identity View offers a straightforward account of some forms of reasoning. Practical reasoning is a species of reasoning toward belief, but a distinctively practical species.

V. Conclusion

The Identity View is well motivated, and can address the counterexamples and abstract challenges to it. It is a compelling conception of intention.

Acknowledgments

Before I met my teachers and peers, Gregory Khasin showed me philosophy and got me hooked. Thanks Grisha.

I learned how to do philosophy from my undergraduate teachers: Christine Korsgaard, Richard Moran, Derek Parfit, Jim Pryor, Thomas Scanlon, and Bharath Vallabha. I learned from them the inescapability of the big problems of philosophy, the importance of thinking clearly and critically about them, and the importance of the history of philosophy for addressing them. In some ways just as important were the members and discussion groups of the Harvard Review of Philosophy, most of all Berislav Marušić and Zoe Vallabha. These were the students who did philosophy even when they weren't required to, and who made it exciting. Quine once said that the mind of the undergraduate is unfathomable; together, these people formed my mind and helped it become, hopefully, something others can fathom.

I am extraordinarily lucky in my graduate mentors. Hubert Dreyfus, Hannah Ginsborg, and R. Jay Wallace have led me in directions I did not expect and would never give up. They have been sharp in their criticisms, and generous with their encouragement and their time. And in a crippling and confusing job market, they have shown me, both with advice and by example, how to become a philosopher while remaining a human being. I am very grateful to them for seeing me through this long and still growing project.

The outstanding philosophical community at UC Berkeley leaves a mark on everyone who comes through it. I absorbed as much as I could of Niko Kolodny, John MacFarlane, and Barry Stroud over many years of courses and interaction. Whatever the topic, they have been constant models, reminding me at every turn what it looks like to do philosophy clearly. I grew up with a large group of exceptional graduate students, many of whom contributed directly to this dissertation; I thank especially Jeremy Carey, Nick French, Tyler Haddow, Jim Hutchinson, Dylan Murray, Antonia Peacocke, Kirsten Pickering, and Janum Sethi. Among graduate students at other institutions, I thank Dylan Bianchi, Benjamin Brast-McKie, Laura Davis, Joshua Eisenthal, Elena Garadja, Joshua Hancox, Nathan Hauthaler, Ulf Hlobil, Adam Marushak, Rebecca Millsop, and Daniel Sharp for their interest, their feedback, and their friendship and belief in me. Among non-philosophers, I give a huge thanks to Myla Green, Ilya Parizhsky, Susana Witte, and Gail Mandella for their enormous friendship and support. And although their numbers are much larger, I have to thank the fantastic undergraduates at Berkeley, whom it has been a joy and a privilege to teach. Special thanks to Yuan Wu, one of the best philosophers I have ever met.

Research on this dissertation was supported by the UC Berkeley Humanities Fellowship for Predoctoral Study; a Dean's Normative Time Fellowship; the Mabelle Macleod Lewis Memorial Fund; and a Berkeley Connect Fellowship. Parts of Chapters 2

and 5 have recently been published as “Moore’s Paradox and Akratic Belief,” © 2014 by *Philosophy and Phenomenological Research*, LLC (forthcoming; published online July 2014), and “A Solution for Buridan’s Ass,” © 2016 by The University of Chicago (published in *Ethics*, January 2016). I am grateful to these fellowships for their support, and to the publishers of *Philosophy and Phenomenological Research* and of *Ethics* for permission to reuse the previously published material.

Most of all, I want to thank the two people who were most with me before, during, and after this dissertation: my mother, Julia Chislenko, and my husband, Charles Goldhaber. My mother cultivated the discernment in my normative beliefs, and the strength of my intentions. My husband became their object. Without them, I could never have connected the two so systematically. Their sharp thinking, extensive feedback, and moral support have led to countless improvements in every chapter. And apart from any details, they have made the hard work so very worth it.

Thank you!!

Chapter 1: The Guise of the Good

Some ancient philosophers saw an essential connection between motivation and evaluation. For them, action was not just a matter of brute pushes or pulls, but an attraction toward something conceived as good. We read in Plato's *Republic* that "Every soul pursues the good and does whatever it does for its sake." Aristotle had a similar view: "It is always the object of desire which produces movement, and this is either good or the apparent good." By Kant's time it was "an old formula of the schools," as he called it, that "We desire nothing except under a conception of the good; we avoid nothing except under a conception of the bad."[1] The idea that we act, or desire, or both, only *sub specie boni*—"under the guise of the good"—is a familiar one in Plato, Aristotle, medieval scholasticism, Kant, and, more recently, Anscombe, Davidson, Korsgaard, Raz, and others. It is one of the central ideas in the history of thinking about action. Some call it the traditional or "scholastic" view.[2]

To many people, this view has come to seem horribly narrow and naïve. But I think that the tradition was on to something very important, and that many of the best resources for defending it have not yet been developed. In this dissertation I defend the traditional view. In the first two sections of this chapter, I introduce some motivations for holding a "guise-of-the-good" view of action (§I), and some challenges such a view must face (§II). But not much progress can be made without noticing that our topic is not one view, but a large family of distinct views. In §III, I use a series of distinctions to introduce and classify the wide variety of guise-of-the-good views. §IV considers strategies for narrowing down one's focus to a particular view or range of views that might be worth defending, and settles on a particularly ambitious view about intention. §V summarizes the remaining chapters of the dissertation, describing how they will address the challenges of §II. As a whole, this introductory chapter lays out the project of defending a guise-of-the-good view.

---

[1] The quotations are from Plato's *Republic*, 505e in the standard Stephanus pagination, in Plato (1997), Aristotle's *On the Soul*, 433a27-29 in the standard Bekker pagination, in Aristotle (1984), and Kant (1997), 5:59 in the standard Akademie pagination. The full formula reads: *nihil appetimus, nisi sub ratione boni; nihil aversamur, nisi sub ratione mali*. For other classic statements, see the *Protagoras*, *Gorgias*, and *Meno* in Plato (1997); the opening lines of Aristotle (1999); Kant (1997, 5:29) and (1998, 4:446-7) in the standard Akademie pagination; Anscombe (1957, 70ff); and Davidson (1980a) and (1980b, esp. 96-102). Recent defenses include Korsgaard (1996, esp. Lecture 3), and (2009); Tenenbaum (2007); Boyle and Lavin (2010); and Raz (2010).

[2] Tenenbaum (2007), for example, consistently uses the term "the scholastic view," echoing Kant's phrase "an old formula of the schools," and emphasizing the view's prominence in medieval scholasticism. I leave a treatment of the history of this view for another occasion.

Such a defense is both deeply motivated and surprisingly complex. It is rarely given a systematic, book-length treatment. My first task is to say why it deserves one.

I. Motivations

The idea that we act "under the guise of the good" has a range of motivations. Most fundamentally, the connection to evaluation has been thought essential to what makes something an action in the first place. When a doctor hits a patient's knee in just the right way, the patient's leg moves. When we walk into a dark room, our pupils dilate. When we eat food, we digest it. But the leg movement, dilation, and digestion are not what we ourselves do. They are not actions, as painting, shopping, or getting married usually are. What makes these actions, some have thought, is at least in part their connection to our own evaluation of our surroundings and of our own activity. We paint because we see it as worthwhile, or see today as a good day for painting. We shop because we believe we should have food and clothing, or because it is good to have some fun on a day off. We get married, ideally, because we love someone and believe he is the right person for us. Seeing what one does as good, and doing it because one sees it this way, can transform a mere bodily movement into an action. If our pupils dilated because we believed it would be good for them to dilate, we might reconsider our denial that dilation is something we do.

There are several variants to the idea that being "under the guise of the good" is a, or the, distinctive characteristic of action. First, there is the thought that the evaluative element is what makes a movement *mine*. My digestion is in one way mine; but we do not say, for example, that I am responsible for it. A second way of putting this thought is that being done under the guise of the good is essential for the attribution of a movement to a *person*. Without someone's evaluation of something as in any way good or worth doing, it can be hard to see why one should think that it is the person herself who is doing it. Third, the notion of action is often thought to be importantly connected to the notion of self-government. An action, one might think, is an instance of self-government—and government not by causal laws of one's own invention, but by one's own conception of what ought to be done. These ideas are interconnected, and there is much to be spelled out in each of them. But they are all ways of seeing evaluation as central to our actions in the first place.

A second, and related, appeal of the guise of the good in the understanding of action is its role in the *explanation* of an action. We might want to know *why* someone is going shopping. The answer might be that her child is growing out of his clothes, and she thinks she should buy him clothes that fit him. At this point one might say: "I see why she thinks she should buy them, but why is she buying them?" But the question seems out of place. The explanation of why she thinks she should buy clothes is itself an explanation of why she buys them. The idea that we act under the guise of the good offers a way to do justice to this sense of explanatory connection. Explaining why someone does something, and explaining why she sees doing it as good, naturally go together. It can seem at best

unnecessary to introduce a further desire to do what one thinks is good to explain why someone shops when she believes she should. A single explanation, we might think, should be able to account for both the evaluation and the action itself.

This second motivation is partly theoretical: the guise of the good offers a kind of simplicity in the explanation of particular actions. But the motivation is not only theoretical simplicity. The unity of explanation can itself seem importantly right. It would be odd, one might think, to have two different explanations that were only contingently related: an explanation of why someone goes shopping, and another explanation of why she thinks it is good to go shopping. To have two separate explanations can seem to miss an important intrinsic connection between an action and the apparent good of performing it.

There is a third motivation which is similarly quasi-theoretical. Belief, many think, is in some significant way governed by truth, or by evidence. This characteristic has been thought to be essential to belief; when we believe, we in some way see our own belief as true, or as supported by the evidence. If we act only under the guise of the good, action (or perhaps intention) is analogous to belief. Both are trying to get things right: belief with respect to the true, action with respect to the good. Or perhaps, one might think: both belief and action or intention take themselves to be in some way justified, or as they ought to be.

I call this third motivation "quasi-theoretical," because it is in part an attractive feature of a theory, on which belief and action or intention can be treated in parallel or analogous ways. The motivation is partly a desire for a unified theory. But as with the second motivation, theoretical considerations may not be as important as the underlying sense that the theory is correct: that we are, at bottom, trying to get things right, both in our understanding of the world and in what we do in it.

These three motivations can lead us to think that there must be something deeply right about the idea that we act under the guise of the good. There are at least three other motivations worth mentioning. These other three are motivations for being *interested* in whether we act under the guise of the good. But they might also provide some support for believing that we do.

A fourth motivation is historical precedent. Many central figures in the history of philosophy have held some version of the idea that we act under the guise of the good. Many others have denied it. This alone is enough to justify being interested in an idea, and in thinking about why so many intelligent people would believe or attack it. It might also be reason to suspect that those who believed it—as well as those who did not—were on to something important. Historical precedent can lead us to ask what insight can be found in those who believed that we act under the guise of the good, and to ask whether their views might be defensible.[3]

[3] For a contrasting view, see Stocker (1979, 739): "Since my main concern is working toward an adequate moral psychology, I shall ignore questions of exactly how and why so many philosophers have held that, of necessity, the good or only the good attracts us."

A fifth motivation for at least considering the guise of the good is its importance in moral philosophy. In the opening of his *Nicomachean Ethics*, Aristotle observes that "Every action and decision seems to seek some good"(1997, 1094a2-3). He then goes on to ask what sort of life is actually a good life. Kant's "supreme principle of morality," in his first formulation of it, tells us to "Act only in accordance with that maxim through which you can at the same time will that it become a universal law"(1998, 4:392 and 4:421). His argument for the validity or bindingness of the principle proceeds by arguing that a will must regard itself as autonomous, in the sense of giving itself a law (1998, 4:446-56). It is not just a matter of convenience that inquiries into the connection between action and evaluation, and into foundational questions in moral philosophy, appear within a single text. The guise of the good may itself have a foundational role to play in moral philosophy and in virtue or moral obligation. If we all pursue what we see as good, a conclusion about the nature of the good will have an inescapable claim on us. There will be no possibility of rejecting the pursuit of the good, and acting in some other way, since that pursuit is ineliminable from action itself. Though this is not a reason to hold a guise-of-the-good view, it has been a source of hope that the view is right. And it *is* a reason to be interested in such a view, and to believe that it is important whether it is true.

Lastly, a guise-of-the-good view can be seen as an extension of ordinary generosity or charity of interpretation. Many of us think it is *kind* to others—and to ourselves—to interpret their motivations in the best light we can. It is a good idea to consider that a compliment might be genuinely appreciative, rather than a piece of sarcasm or self-serving flattery. We can try to remember that a thief, or an addict, or an obnoxious colleague is doing the best he can under difficult circumstances. A guise-of-the-good view is generosity of interpretation on a grand scale, systematized into a view about acting beings. It says that even when we are conflicted, confused, or exhausted, we pursue what we see as good. This might again not be a reason to believe that a guise-of-the-good view is true. But it can lead us to ask whether such a generous view of human nature can be defensible.[4]

All six of these considerations—the nature of action, unity of explanation, parallels with belief, historical precedent, moral implications, and generosity of interpretation—are motivations for a systematic consideration of the viability of a guise-of-the-good view. As we will soon see, the project of such a consideration is very large. One way I will keep it manageable is by limiting my consideration of its broader implications. I leave open the possibility that some of these motivations are themselves illegitimate, or that other, competing views can be motivated in similar ways. My task here is a consideration of the guise of the good itself. But in pursuing that task, it is worth keeping in mind its connection to these many larger issues.

---

[4] There is, of course, a limit to how generous a guise-of-the-good view can be. Even if we, in a sense, always mean well, it can still be appropriate to remember the saying: "The road to hell is paved with good intentions." And on a guise-of-the-good view, the worst and most malicious acts are committed with a view to an apparent good. So it is worth remembering that the view offers generosity in a particular respect. It says that we always pursue what we see as good, rather than remaining indifferent to it or pursuing the bad as such.

II. Challenges

At least on the face of it, it can seem obvious that we often do and want what we do *not* see as good. In one kind of example, we act akratically, or against our better judgment, doing what we believe we should not do. We take an extra helping of ice cream, or stay up too late, or insult or hit a child, though we believe we should not. Our actions and desires on the one hand, and our evaluations on the other, seem to be in conflict. It seems undeniable that akratic action is possible or even common. But then how can guise-of-the-good views account for such widespread conflict between our actions and what we do conceive of "under the guise of the good"?

In other cases, we seem to act without having an evaluation at all one way or the other. Buridan's Ass starved when faced with two identical and equidistant bales of hay. We often face two identical and equidistant pieces of food, and we do not starve. We manage to take one, apparently without singling it out in evaluation. At other times, we choose to go to a museum rather than a concert, or care for a sick relative rather than fight for a worthy cause, not because we conclude that it is the better option, but because we are unable to come to a conclusion about which one is better. In each case, we can believe what we do to be good overall, but it seems we at least do not believe it to be better, or to be what we ought to do. In other apparent examples of lack of evaluation, the point may be more general. Some creatures who act may not be old enough, or cognitively sophisticated enough, to have or use evaluative concepts. Young children, non-human animals, and human adults with severe enough impairments can seem able to act without even having the concept of goodness.

Just as we seem able to act without evaluation, we seem able to evaluate without a corresponding action. We might believe it would be good to get out of bed, but not move an inch. Many of us agree that we should donate to charity, but cannot claim that we do, or that we intend to, or even that we have the inclination. Here it again seems unlikely that a kind of evaluation could be closely tied to action, since we seem able to have the evaluation without acting on it.

These are three kinds of striking divergence between action and evaluation. If we think of action, intention, and desire as falling under the general heading of "motivation," we can summarize them succinctly. They are examples of evaluation and motivation in conflict, motivation without evaluation, and evaluation without motivation.[5] They all seem to be common and easily recognizable cases. They can make a guise-of-the-good view seem hopelessly narrow, naïve, and most importantly, false. Faced with these examples, a guise-of-the-good view no longer seems to capture anything about action or desire in general. It seems to describe only, as David Velleman put it, "a particular species of agent, and a particularly bland species of agent, at that"(2000, 99). If this is right, a guise-of-the-

[5] For another defense of a guise-of-the-good view that distinguishes and addresses these three kinds of counterexample, see Tenenbaum (2007), Chapters 6-8.

good view cannot capture anything general about action or its explanation, or offer a general parallel with belief or ethical implications of the kind I described. And clinging to a guise-of-the-good view would be an extension, not of ordinary, reasonable generosity of interpretation, but of a blind optimism that fails to see the person it is trying to be kind to. It would amount to, as Stocker put it, an "unjustifiable optimism or complacency"(1979, 749) with respect to a view that everyday examples show to be "clearly and simply false"(1979, 740). To understand action is, it seems, in part to understand all the ways we can fail to pursue what we see as good.

Apart from particular counterexamples, there are also more abstract challenges to a guise-of-the-good view. Action, for example, seems to be under our voluntary control. Evaluation does not; so if action required evaluation, it seems, it could not be voluntary in the way that it is. Secondly, evaluative judgment and belief seem aimed at accuracy, and to change when we see that they are incorrect. Motivational states such as intention and desire seem directed not at matching how the world actually is, but at changing it to match them. They seem to have the opposite 'direction of fit'. They then seem to be a different kind of state from evaluative judgment or belief, and it can seem unlikely that they require any kind of evaluation. Third, evaluation and motivation can seem to differ with respect to reasoning. 'Theoretical' and 'practical' reasoning can seem fundamentally different in the questions they address, the procedures they use, or their interest in conclusions that can be true or false.

These are varied and powerful challenges. Few, if any, of them have an obvious and immediately convincing answer. Together, they are formidable. If one hears that we can only do what we conceive of as good, it is natural to ask: but then how can we do what we believe is not good? What can this view say about Buridan's Ass, or about young children? How can our actions still be voluntary? Answering these and other challenges will be the main task of the chapters that follow. But first, it is worth being more precise about the kind of view that is at issue.

III. Parameters

So far I have spoken loosely, without sharply distinguishing alternative views. But we can pause to see what kinds of guise-of-the-good view are possible. With the general motivations for these views in mind, a more detailed introduction of the views themselves is a second step toward seeing what view might be worth defending. We can ask several questions, each corresponding to a set of distinctions.

First, one might ask: what is the central topic here? Is it action or desire? Is the idea that *wanting* to buy clothes requires seeing the buying as good in some way, or rather that actually buying them does? The answer, of course, is that both views can and have been held. Aristotle combined them when he wrote: "It is always the object of desire which produces movement, and this is either good or the apparent good"(1984, 433a27-29). If

this view is right, both desire and movement require the object in question to be or at least appear good.

Intention, too, can be thought to be essentially evaluative. We might think that one can only intend to do what one sees as in some way good, or worth doing. We can thus talk of a family of views: views that describe action, intention, and/or desire as something had or performed "under the guise of the good." We can call such views "guise-of-the-good views" in each case. As I have described it, the core idea of a guise-of-the-good view is the attribution of an evaluation. The idea is that action, intention, and/or desire involve an evaluative component: in the classic versions I mentioned, a conception or appearance of goodness.

Second, what does this "involving" amount to? It might be that, as Davidson (1980b, 99) put it, "the intention simply is an all-out judgment." Perhaps, in other words, an action, intention, or desire is itself a way of seeing something as good.[6] Or instead, the idea might be that there is a necessary connection with evaluation, so that, for example, we cannot desire anything we do not see as good. Or maybe there is a weaker connection—maybe we must only, for example, be *able* to see what we do as good, even though we sometimes do not. The relation to evaluation can be identity, necessity, or something less. Which of these one has in mind again affects the sort of guise-of-the-good view one holds.

Third, what sort of evaluation is involved? Though it is tempting to speak vaguely of "seeing" something as good, the "seeing" is not literally a visual perception. It might be more accurate to say we must *believe* what we do to be good. Or maybe the evaluation is a kind of judgment, distinct from belief. Or maybe it is neither of these. Tenenbaum (2007) argues that "desires are best conceived of as appearances of the good"(2007, 17). The way an argument can seem valid, or an action can seem wrong, Tenenbaum thinks, desire is itself something's *seeming* good to us. Once again, all of these views are possible: one can understand the "guise" or "conception" of the good as a belief, a judgment, or an appearance.

Fourth, we can also ask: What is it that is conceived to be good? Is it the action that one performs, or wants or intends to perform? Or is it that we act "for the sake of the good," in the sense that we always pursue *ends* that we believe are good? Or maybe what we must see as good is the outcome of the action: we always pursue or desire an apparently good state of affairs. There is room to be more precise even about the object of evaluation.

Fifth, although I have been talking in terms of goodness, 'good' is not the only available evaluative concept. An ambitious guise-of-the-good view can insist that we do only what we believe we *ought* to do. Even with respect to goodness, further distinctions

[6] Davidson also suggests such a view about intentional action: "In the case of intentional action, at least when the action is of brief duration, nothing seems to stand in the way of an Aristotelian identification of the action with a judgement of a certain kind—an all-out, unconditional judgement that the action is desirable (or has some other positive characteristic)"(1980b, 99). For an analogous view about desire and intention, see Tenenbaum (2007), for whom "Desires are *appearances* of the good"(27), and intentions are "unconditional evaluative judgments that either are embodied in or precede action. Just as in the case of desire, we should not think of the judgment as something other than the intention"(66).

can be made. Some of us might think that we can only do what we believe is good overall, as abstaining from dessert often is. For others, this might seem to claim too much, but to still carry a grain of truth: it might be that we must at least see *something* good in what we do—some good characteristic in the action, end, or outcome, even if it is only the pleasant sweetness of an extra piece of food.

We can call each of these five dimensions of variation a "parameter." The parameters I mentioned are: (1) the target of the view, or what the guise-of-the-good view is about; (2) the type of connection with evaluation; (3) the nature of the evaluation; (4) the subject of the evaluation, or what the evaluation is about; and (5) the central notion used in the evaluation. Making a set of selections within the parameters formulates a guise-of-the-good view. For example, one might say: "Action requires a belief that the action is good." Or: "Action requires a capacity to judge that its outcome will be in some way good." Or: "An intention to act is a belief that one ought to perform that action." Or: "A desire to do something is an appearance that the action is good."

The resulting range of views can be summarized in the following template:

The Guise-of-the-Good Template

> An (action / intention / desire) (is / requires / requires a capacity to have) a (belief / judgment / appearance) that the (action / end / outcome) is (what one ought to do or bring about / good / in some way good).

Multiplying the numbers of options within each parameter in the template gives us $3^5 = 243$ logically possible guise-of-the-good views.

Even the Guise-of-the-Good Template is far from exhaustive. It is a starting point for articulating a view, and many other variations are possible. We can even classify some types of variation. There can be:

1. *Further specification*

A guise-of-the-good view can make a further specification in one or more parameters. For example, take the view: "Action requires a belief that the action is good." If I tell a lie, on this view, I must believe that telling the lie is good. Though an element in each parameter has been specified, each one can be made more specific. (1) Maybe only a particular kind of action requires such a belief: only intentional action, for example, or only human action, or only divine action. (2) The requirement can be understood as a conceptual, metaphysical, or even physical necessity. Perhaps it is part of the nature of action that it be accompanied by a belief that the action is good, even though it is not part of our notion of action. (3) I might have to believe in the lie's goodness consciously, or with a certain level of credence, or at least partially or implicitly, or in a way that constitutes knowledge. (4) What is believed good could be the individual lie I will tell, or lying as a type of action, or lying in particular circumstances. (5) Lastly, I can believe my

lying to be good morally, from an impartial point of view, for me, or in some other way. These specifications can be combined. One further specified view would be: "Intentional action metaphysically requires a conscious belief that acting this way in relevantly similar circumstances is good for the acting person." This is just one example of further specification in each of the five parameters. We can imagine indefinitely many more.

2. Addition

Instead of making an existing element in the template more specific, one can add a possibility that I did not include at all. Examples can again be found for all five parameters. (1) Rather than action, intention, or desire, we might be interested in willing, or in habit. If the guise of the good has isolated exceptions, perhaps an action can still become habitual only if one believes it is a good action to perform.[7] (2) One might think that, for example, action must be *motivated by* a belief that the action is good. (3) Action can be thought to require a *feeling of conviction* that the action is good—a feeling that is not itself a belief, judgment, or desire, and might not require or be required by any of those. (4) Action (or, more plausibly, intention) might require a belief that the *intention* is a good one to have. (5) Lastly, the central evaluative notion can be thought to be a different one. A guise-of-the-good view can be formulated in terms of *appropriateness*, or *desirability*, or what we have *reason* to do. Each of these additions can be subject to further specification.

Apart from additions within each parameter, we can add an entire parameter that has so far been left out. For example, one might want to add a parameter specifying *who* has the belief, judgment, or appearance. I have assumed it is the same person who performs the action or has the intention or desire. But the link to evaluation can be understood differently. On one conceivable variant, we can only act if God believes our action is good. On another, action requires some part of us—some cognitive subsystem, or some part of the psyche—to believe or judge that the action is good, even if the person as a whole does not.

Some of these additions might seem farfetched, or less important to consider. I hope they do, since I tried to include the most important variants in the template. But they are not clearly insignificant, and there are likely to be others. I make no claim to have exhausted the range of guise-of-the-good views that have actually been held, or that are worth considering.

[7] Guise-of-the-good views with respect to willing are common in Kantian moral philosophy. See, for example, Korsgaard's insistence that "You must will your maxims as universal laws"(2009, 72). For another possibility, see Setiya (2007), who argues mainly against the idea that "*reasons for action* must be seen under the guise of the good"(21).

3. Interaction

A guise-of-the-good view can interact with a person's views about other, closely related topics. For example, some people think of action as falling under a broader category of activity or behavior.[8] On such a view, another possibility can be added to the first parameter in the template, and another set of guise-of-the-good views can be formulated about activity or behavior in general. If there is no distinct category of activity or behavior, there is no further distinction or addition to be made here. Similarly, one can disagree about the relation of what we ought to do with what we have most reason to do. If these notions are distinct, we can distinguish guise-of-the-good views corresponding to them. If they are not, we cannot. Even for a particular notion such as the notion of goodness, or 'ought', or 'intention', there can be disagreement about whether the notion is ambiguous. Ambiguity again multiplies possible views. In general, a conception of the nature of each element in each parameter, and of the distinctions and relations between the elements, can have consequences for a conception of the variety of conceivable guise-of-the-good views. The range of conceivable guise-of-the-good views can thus be expanded to accommodate the range of relevant views on related topics.

Though the Guise-of-the-Good Template includes 243 views, the further specifications, additions, and interactions could easily reach the tens of thousands. This overabundance is a mixed blessing. On the one hand, it raises many interesting questions; it brings out the intricacy of the topic; and it helps locate any particular view in a landscape of alternatives. On the other hand, it would be extremely cumbersome to constantly distinguish all of these possibilities. Even in describing each parameter, I sometimes ignored some of the others. I have chosen to emphasize the variety of views in some detail in this section, rather than reiterate it more crudely at every step elsewhere. But it is worth keeping this variety in mind. It is worth asking which views are being ignored, and which views are being conflated with each other.

More immediately, the variety of views creates a problem of selection. In considering guise-of-the-good views, it becomes difficult to know where to start. Which views are worth even trying to defend? My larger goal is to find a view that claims something important, unobvious, and true. In the next section, I ask how one can narrow one's focus to one or at most a few of the many alternatives.

IV. Strategies

The motivations, challenges, and parameters for guise-of-the-good views all make the task of a defense daunting. Which motivations are worth doing justice to, and how? How can one answer all of these challenges? Which of the thousands of conceivable guise-

[8] See, for example, (Frankfurt 1988b, 58) and (Velleman 2000b, 2).

of-the-good views should we start with? People sometimes say that a full consideration of an issue would be a book-length project. But there is no hope of covering so much ground in a single book. I will have to be selective in the choice of views to consider.

One natural goal of a selection is simplicity. There is a conceivable view that can be called the 0.7-Credence-Health View. On this view, deliberatively endorsed intentional action metaphysically requires a conscious belief of credence 0.7 or higher that pursuing the action's goal in relevantly similar circumstances is good for the health of the acting person. Among other failings, such a view seems unnecessarily complicated. Of course, one must be careful not to oversimplify, either, since the right view might be complex. But it is useful to start with relatively simple views, to see what complications might be needed.

A second guiding consideration is historical precedent. It can be useful to begin with views that have been held, or at least attacked, by influential figures. I will not assume that any historical figure is infallible; but I do assume that widely held views are worth considering, and might be on to something important. If many influential philosophers had held the 0.7-Credence-Health View, or argued against it, it would call for more serious attention.

A third consideration is doing justice to the motivations for guise-of-the-good views in general. The 0.7-Credence-Health View is not obviously attractive in any of the ways I considered in §I, and it is unclear what other motivations it could have. Being obviously well motivated is not indispensable, since some motivations can be unobvious. But in general, we should be more interested in views when we can see why they are worth taking seriously.

A fourth consideration is ambition. On the one hand, some logically possible guise-of-the-good views can make such extreme claims that there is no hope of defending them. I will not consider the view that animal action requires a belief that all conscious beings should perform that action under all circumstances. It is hard to imagine why one would ever believe this. The 0.7-Credence-Health View, too, is in one way very ambitious. It claims that all deliberatively endorsed intentional action requires a belief in a pursuit's goodness *for one's health*. Here there are powerful counterexamples. People sometimes take on grave health risks, or even gives up their lives, for the good of their family or country. Requiring a belief that a pursuit is good for one's own health thus seems overly ambitious and even extreme. On the other hand, many guise-of-the-good views have been thought to make extreme claims that there is no hope of defending. It is worth being cautious in deciding in advance that a view is overly ambitious.

A guise-of-the-good view can also be underambitious. For example, consider a kind of simple "noncognitivist" view about evaluation: that to conceive of something as good is not to have a belief about it, but simply to want it. On such a view, we can still say that wanting requires conceiving as good. This would even be true. But it would be trivially true. It is then not worth considering for long, or defending in detail. Other views can be underambitious without being tautologous. One might think that deliberatively endorsed intentional action requires a belief that there is something good about performing

the action. Though this is a guise-of-the-good view, it requires very little. It is easy to see *something* good in an action. Overeating can be pleasant; child abuse releases anger; genocide reduces overpopulation. It is hard to think of even one conceivable action without a single good characteristic. Here there is room for more substantive debate than there is about the noncognitivist view. But even if such a view is true, it does not tell us much that is interesting about ourselves. When considering guise-of-the-good views, it is worth looking for the views that are ambitious enough to be interesting, but modest enough to be defensible.[9]

These strategies are largely methodological. In part, they are answers to the question of which views to consider first. I will follow all of them. I will begin with a range of views that have been both motivated and widely held. Of these, we can start with

---

[9] Both defenders and deniers of guise-of-the-good views have drawn attention to unproblematic versions of them. Tenenbaum (2007, 2), beginning a book-length defense, notes that "one could "define the 'good' so broadly that it would end up simply being another word for 'possible object of desire'." Setiya (2007, 62) writes, more generally, that "It is worth being careful…about *trivialized* versions of the guise of the good." Presumably these trivialized versions are true, though they may not be worth defending.

Both defenders and deniers of guise-of-the-good views have also endorsed modest but non-trivial views. Velleman (2000a), in resisting guise-of-the-good views, writes: "I am not opposed to describing desire as the attitude of regarding something as good, so long as this description is taken merely to express the attitude's direction of fit." Audi (1979, 194), though resisting a guise-of-the-good view, writes: "There is at least a *tendency* for one's motivation to accord with one's practical judgments, and nothing I have said precludes some kind of 'non-contingent' relation between the two." Raz (2010) defends a view that is in danger of being underambitious. As he puts it, he "does not assume that agents capable of intentional action must have the concepts used in stating the [guise-of-the-good] Thesis…, nor…that they believe that these concepts apply to each of their intentional actions. It assumes [only] that they have a belief that…can be truly characterized as a belief that the action has a good-making property"(2010, 114). Raz requires only a belief in a good-making property, rather than in overall goodness; and he does not even require possession of the concept 'good'. Apart from doubts about the coherence of this view, one might doubt that it is ambitious enough to be interesting.

Despite attacking guise-of-the-good views, Stocker (1979, 740-1) insists that even some of the more ambitious ones are uninteresting:

> How are we to understand the relation between the good and attraction? It is too weak to require only that the attractive act or act-feature is, e.g., (believed) good in some respect or over-all or even best. For unless such acts or features are (believed) absolutely good—i.e., with no aspects that are (believed) bad or neutral in any respect—they can attract because or only because they are (believed) bad or neutral in some respect or other. Thus this requirement does not give an interesting version of the thesis that the good always attracts or that only the good attracts, that we always act *sub specie boni*. These require that the (believed) goodness or the (believed) good qua good is somehow essential to the attraction: e.g., that acts or features attract because or only because they are (believed) good. It remains problematic exactly how to specify this requirement.

For Stocker, the interestingly controversial idea is not that some evaluative attitude is present, but that it plays some 'attracting' or motivational role.

simpler views, and see what must be made more complex. And we can start with ambitious views, and see what needs to be weakened. I will try to begin with the simplest and most ambitious of the influential and well-motivated guise-of-the-good views. Although desire is an important topic in its own right, I will leave it aside, and focus on action. More specifically, my focus will be on intentional action, in the sense of action that is, at least under some description, intended.[10] The movements of a cat or a human baby are naturally thought of as actions, as are the fidgeting, humming, or pacing of an adult human—even when they are not actions that their doer intends to perform. But my focus here will be on the actions that we do perform intentionally. I take this narrower focus, partly because of the centrality of intentional action in a distinctively human life; partly to keep the size of the inquiry manageable; and partly because, as will emerge in the chapters that follow, I believe it is here that an especially ambitious guise-of-the-good view can be defended. I will neither defend nor deny a guise-of-the-good view about desire, or about non-intentional action.

In considering guise-of-the-good views about intentional action, it is useful to think directly about intention. A conclusion about intention will have direct implications for a view about the actions we intend; if intention must be in some sense "under the guise of the good," then so must intentional action. But a guise-of-the-good view about intention has a broader range, since it applies also to intentions that are not carried out. And guise-of-the-good views about intention can sometimes be simpler to consider; even if intention is itself a belief about goodness, it is hard to see what it would mean for an intentional action to itself be a belief of any kind. From here on, I will focus mainly on intention, though I will of course consider many kinds of intentional action along the way.

What guise-of-the-good views about intention are worth considering? Recall the Guise-of-the-Good Template: "An (action / intention / desire) (is / requires / requires a capacity to have) a (belief / judgment / appearance) that the (action / end / outcome) is (what one ought to do or bring about / good / in some way good)." One especially ambitious view of intention would hold that intention is itself a belief that one ought to perform the action. Put more formally, letting "A" stand for a person[11] and "*x*" for an action, this view can be called

The Identity View: A's intention to *x* is a belief that A ought to *x*.

---

[10] The focus on intentional action has become common in the contemporary literature. Davidson (1980a) focuses most centrally on the view that "in so far as a person acts intentionally he acts…in the light of some imagined good"(1980a, 22; cf. 1980b). Velleman (2000a) centers an essay titled "The Guise of the Good" on a family of views focused on intentional action: "Intentional actions are aimed at the good"(100); "An agent…does nothing intentionally unless he regards it or its consequences as desirable"(99); "All…intentional actions are…directed at outcomes regarded…under the guise of the good"(99); "[An] agent is…capable of intentional action…only by virtue of being a pursuer of value"(99).

[11] Here I do not mean "person" in the biological sense of a member of the human species, but any being capable of belief and intention. On non-biological senses of "person", see Frankfurt (1988a, 11-12).

On the Identity View, the intention *is* a belief, rather than merely requiring one. It is a *belief*, rather than a mere appearance or inclination to believe. It is a belief that the particular *action*, rather than, more broadly, the pursued end or outcome, is what A ought to do. It is a normative belief: that is, a belief that A *ought* to *x*, rather than that *x*'ing would be one good alternative among others. And the Identity View makes clear that A's belief applies to herself; she believes that she herself ought to x, rather than that, for example, someone should.[12]

The Identity View can be weakened in various ways. Without a claim of identity, it becomes a less ambitious view, which can be called

> *The Normative Belief Requirement*: If A intends to *x*, A must believe she ought to *x*.

This requirement is entailed by the Identity View. If A's intention to *x* is itself a belief that she ought to *x*, then A cannot have the intention without the belief. But the Normative Belief Requirement says nothing further about the nature of intention, and leaves the type and explanation of the necessity unspecified.

The Normative Belief Requirement can be weakened further. Without requiring a belief, the view could become

> *The Normative Appearance Requirement*: If A intends to *x*, it must appear to A that she ought to *x*.

This weaker requirement is more intuitively acceptable; it is easier to believe that if we intend to do something, it must at least in some way seem to us that we ought to do it. Without the 'ought', the view could become

> *The Appearance-of-Goodness Requirement*: If A intends to *x*, it must appear to A that it is good to *x*.

This requirement again claims less than the previous one. And using the Guise-of-the-Good Template, we can find even weaker views. "Good" can be replaced with "in some way good." Requirement can be replaced with a requirement of a capacity. "x" can be replaced with "something." We would then have

> *The Evaluative Capacity Requirement*: If A has an intention, A must have the capacity to have it seem to her that something is in some way good.

[12] Thus not all normative beliefs are intentions, even on the Identity View. I discuss the relevant kind of normative belief in more detail in Chapter 7.

On the Evaluative Capacity Requirement, a being with intentions must at least be able to have things appear good to her. This requirement still claims a kind of necessary connection with evaluative states. It is not tautologous, or in any other sense completely trivial; even if it is true, it is interesting that it would be true. But this requirement is hardly ambitious. Though it is still in the family of guise-of-the-good views, it is much less likely to be called ridiculous, or obviously false. On the contrary, it can seem too watered down to be interestingly controversial.

The mere possibility of very modest views already has one implication: It would be too quick to proclaim that all guise-of-the-good views are obviously false. There are too many such views, and too many and too varied potential weakenings of them, for anyone to reasonably be sure that all of them are misguided. Instead of dismissing all guise-of-the-good views offhand, we can ask which ones could be true.

There is now a natural starting point. To begin with the simplest and most ambitious view, and see what must be weakened or made complex, one can look to the Identity View itself. The Identity View is very simple and very ambitious. If true, it does a great deal of justice to the motivations for guise-of-the-good views. It attributes an evaluative element to intention and intentional action, draws more than a close parallel between intention and belief, and sees beings who act intentionally as essentially trying to get things right. It is extreme enough to have little historical precedent.[13] But it is an ambitious version of a family of central guise-of-the-good views, which draw an essential connection between motivating and evaluative states. For all these reasons, it is a useful place to begin.

My strategy, in other words, is to begin with a simple and ambitious guise-of-the-good view, and consider how well it meets the challenges that can be raised against it. Here there is another reason to begin with the Identity View. Unlike some more modest guise-of-the-good views, the Identity View faces all of the counterexamples and abstract challenges I mentioned. This can be thought to be a disadvantage for the view itself. But it will allow us to consider all of the leading challenges to guise-of-the-good views, and to see what it takes to address them. This is an added reason to begin with the Identity View, and it sets the agenda for the chapters that follow.

V. Summary of the Dissertation

So far, I have said why the Identity View is a good starting point in considering guise-of-the-good views of intentional action. But in fact, I believe that a more detailed consideration of the Identity View shows it to be defensible. Rather than say how it must be weakened, I am going to defend it. I think it can meet all of the challenges to guise-of-the-good-views, and offer a compelling general theory of intention. Although the strategy was to begin with a simple and ambitious view, I believe the conclusion this strategy

[13] It does have some; see Kant (1997 and 1998), Davidson (1980b), Korsgaard (1996 and 2009), and Tenenbaum (2007), mentioned above.

should lead us to is that the view should be accepted. This dissertation is a defense of the Identity View.

As with other views, not only weakenings, but also further specifications, additions, and interactions can be imagined. The Identity View can be held about a particular type of intention, for example, or with "ought" used in a particular moral, prudential, or other sense. I will have little to say about the nature of the normative "ought." I do have in mind genuine normative belief, rather than belief that some norm of etiquette applies, as we might express in saying that one ought to eat salad with the smaller fork. But I will treat "ought" as a primitive concept, and use "should" interchangeably with it. If it is to be further specified, the general form of the defense I give can be applied to a more particular Identity View as well. Or at least, so I hope. In what follows, I will offer a general defense of the Identity View, which may well apply to various weaker guise-of-the-good views. Potential implications to a range of more modest views is one motivation for considering a single, ambitious one. But I will leave these more modest views aside, since I believe a retreat to them is unnecessary.[14] And I will mostly leave aside further conceivable variants of the Identity View, to focus on defending the view itself.

The next five chapters address central counterexamples to guise-of-the-good views, and begin to develop the general conception of intention expressed by the Identity View. In Chapter 7, having worked out some of the details and implications of the Identity View, I turn to that general conception by considering the more abstract challenges, and arguing that the Identity View can meet them as well. I'll now briefly summarize these chapters; for a section-by-section summary, see the Outline at the end of the Table of Contents.

The aim in each chapter is partly defensive. Each of Chapters 2-6 considers what can seem to be a knockdown objection to guise-of-the-good views, and in particular to the Identity View. One of the aims of Chapters 2-6 is to show that there are no knockdown objections, or even any seriously damaging counterexamples, to the Identity View. But none of these chapters is entirely defensive. A guiding idea about the nature of intention should be able to shed light on the details of everyday life. In Chapters 2-6, I try to show how the Identity View can improve our understanding of the very cases that have been thought to be problematic for it.

I begin in Chapters 2-3 with akratic belief: *believing* what one believes one ought not believe. If it is possible to have intentions we believe we ought not have, and those intentions are themselves beliefs, we can expect that it is also possible to have beliefs one believes one ought not have. I begin with akratic beliefs, first, because they have been widely thought to be impossible, and a commitment to their possibility can be thought to

[14] Since the Identity View entails the Normative Belief Requirement, a defense of the first is also a defense of the second. Some of the following chapters will apply specifically to the Identity View, and if those are mistaken, the others remain as a defense of the Normative Belief Requirement. But without the Identity View, there is the additional problem of why there should be any necessary connection between intention and normative belief. I will not address that problem; indeed, it is one of the motivations for looking to a defense of the Identity View. Without the identity, the necessary connection itself becomes mysterious.

be decisive against the Identity View. Second, I consider akratic beliefs because a consideration of belief outside the context of intention sets the stage for the discussion of belief in the chapters that follow. In Chapter 2, I consider arguments against the possibility of akratic belief, and argue that they have no independent force, apart from expressing an underlying puzzlement about how akratic belief is possible. In Chapter 3, I address that puzzlement by developing a conception of akratic belief. I argue that, by distinguishing several key characteristics of belief, we can come to understand how belief can be akratic.

Chapter 4 turns to akratic action: intentionally doing what one believes one ought not do. In this case, the problem is that such actions do seem possible. Though the Identity View can seem to rule out the possibility of akratic action, I will argue that it instead leads to a conception of akratic action as a manifestation of conflict between normative beliefs. The latter part of Chapter 4 defends that conception.

Chapter 5 considers apparent cases of lack of evaluation, focusing on Buridan's Ass and inability to compare alternatives. In these cases, we seem to have an intention without a normative belief: of two identical pieces of food, for example, we intend to take the one on the left without believing we ought to take that one. I argue that, when we form intentions in such cases, it is normally by intending to act *non*-intentionally to determine which alternative we will take, and then, once the tie is broken, intending to take that one. There is thus nothing we intend to do in these cases except what we believe we ought to do.

Chapter 6 considers lack of motivation: for example, to get out of bed, make a phone call, or donate to charity. Here it seems there is a normative belief, of the kind that constitutes intention on the Identity View, but no intention. I argue that, when they are not akratic, such cases can be understood as a kind of psychological fatigue, in which we do intend to get out of bed, but fail to execute our intention. It is thus unnecessary, and misleading, to deny that we have intentions in these cases.

Chapter 7 turns to what I have been calling the abstract challenges to guise-of-the-good views. Here I consider direction of fit, voluntary control, and distinctions between theoretical and practical reasoning. I argue that there is independent reason to think that intentions and normative beliefs each have two directions of fit, aiming to match what one ought to do, but change the world to make it as we intend or believe it ought to be; that, in the sense in which beliefs are not voluntary, intentions are not voluntary either; and that the normative beliefs that constitute intentions on the Identity View are themselves, in the relevant sense, practical states, reachable by reasoning that can be properly considered practical. I conclude that the Identity View is a well motivated, surprisingly ambitious, defensible, and illuminating conception of intention and intentional action. We should, I believe, accept that it is true. To intend to do something is to believe that one should do it.

Chapter 2: Believing Against One's Better Judgment, I: Impossibility Arguments

I. The Problem

Before considering akratic action and intention, it will help to look at a parallel phenomenon in belief. Believing against one's own better judgment – 'epistemic' or 'doxastic' *akrasia*, or *akrasia* in belief – is having a belief one believes one should not have.[15] To put it more formally: A's belief that *p* is akratic if and only if A believes that A should not believe that *p*. It is natural to wonder how we can have such beliefs. If we believe we should not have some belief, do we not give it up? The phenomenon akratic belief is itself puzzling and paradoxical. Like *akrasia* in action and intention, it has been thought to be impossible.

To bring out the problem, consider

> *Fear of Flying*: Matt is extremely afraid of flying. When professional obligations require him to travel (even thousands of miles), he either drives or takes a train. He does not travel overseas. When his friends and loved ones travel by air, he obsessively checks the status of their flights online, and calls them as soon as possible after landing to make sure that they're all right. When asked about all this behavior, he doesn't defend it. Instead, he says things like the following: "Of course the evidence shows that flying is not particularly dangerous—certainly less dangerous than driving comparable distances, but I just can't shake the belief that if I fly, my plane will crash and I will die. What's holding it up there anyway?"[16]

According to Matt, he is not only *acting* in a way he does not defend. He "can't shake" the *belief* that his plane will crash—even though he thinks the evidence shows that flying is not dangerous. He seems to be holding a belief against his own better judgment.

Any example of akratic belief can seem puzzling, and this one is no exception. To be akratic, Matt has to have both the belief that his plane will crash, and the belief that he

---

[15] From here on I will avoid most of the roughly synonymous terms I mention here. "Epistemic *akrasia*," though common in the literature on akratic belief, is misleading, since 'epistemic' suggests that the topic is akratic knowledge rather than belief. "Doxastic" is closer, suggesting that the *akrasia* pertains to belief. But it adds nothing helpful to talking simply of akratic belief, and does not as clearly exclude akratic suspension of belief, which is not my topic here. The phrase "against one's better judgment" is worth keeping in mind since it is so widely used, but it brings a danger of uncritically identifying judgment with belief, and raises questions about the sense in which the better judgment is "better." To avoid these complications, I talk simply of akratic belief, and beliefs one believes one should not have.

[16] I borrow this example from Greco (2014, 202). See also Shah and Velleman (2005, 507-8).

should not have that belief. Does he really have both? It seems not only funny but revealing that he ends by asking: "What's holding it up there anyway?" According to the evidence as Matt describes it earlier, there is plenty to keep the plane in the air, enough to make flying "not particularly dangerous." But when he blurts out: "What's holding it up there anyway?", is he not revealing a different view of the evidence? Or if he does believe the evidence, isn't his "belief" that the plane will crash more a recurring fearful thought, rather than a genuine belief? It is hard to make out clearly both the belief and the 'better judgment' that makes it akratic.

Descartes and Hume gave voice to the sense that the conclusions of critical reflection are hard to hold on to. When Descartes wrote that an omnipotent God, or a malicious demon, or fate could deceive him in his most basic beliefs, he went on:

> I…am finally compelled to admit that there is not one of my former beliefs about which a doubt may not properly be raised…. I must withhold my assent from these former beliefs just as carefully as I would from obvious falsehoods, if I want to discover any certainty…. But it is not enough merely to have noticed this; I must make an effort to remember it. My habitual opinions keep coming back, and, despite my wishes, they capture my belief, which is as it were bound over to them as a result of long occupation and the law of custom…. I shall…resolutely guard against assenting to any falsehoods….But this is an arduous undertaking, and a kind of laziness brings me back to normal life.[17]

Hume similarly wrote:

> I am ready to reject all belief and reasoning, and can look upon no opinion even as more probable or likely than another…. Since reason is incapable of dispelling these clouds, nature herself suffices to that purpose, and cures me of this philosophical melancholy and delirium, either by relaxing this bent of mind, or by some avocation, and lively impression of my senses, which obliterate all these chimeras. I dine, I play a game of back-gammon, I converse, and am merry with my friends; and when after three or four hour's amusement, I wou'd return to these speculations, they appear so cold, and strain'd, and ridiculous, that I cannot find in my heart to enter into them any farther.[18]

Both Descartes and Hume here reflectively reject their ordinary beliefs. As Descartes puts it, "I must withhold my assent from these former beliefs"; Hume too is "ready to reject" them, and "can look upon no opinion even as more probable…than another." But the ordinary beliefs resist reflective scrutiny, and seem to have a force of their own. Descartes

[17] Descartes (1641), VII.21-22.

[18] Hume (1739), I.4.7. Both of these passages come at the culmination of a powerful skeptical opening. The passage from Descartes comes from near the end of his first Meditation; Hume's is near the end of the first Book of the *Treatise of Human Nature*.

describes habitual opinions "coming back" and "captur[ing]" his belief despite his wishes; Hume writes that "reason is incapable" of resolving his predicament, but, as in Descartes, occupation and custom loosen the results of his inquiry until he can no longer take them seriously.

Descartes and Hume both describe a change over time. Our better judgment rejects our "former" beliefs, but our "habitual opinions keep coming back"; laziness and distraction bring us back to normal life. And this can seem to speak *against* the possibility of akratic belief, rather than for it. Forgetting, wavering, or changing our minds may seem to be as close as we can get to believing something against our own, current, judgment about what we should be believe. The question is: is there more to what Descartes and Hume describe than a mere forgetting or change of mind? Can we have a belief even as we hold on to the belief that we should not have it?

I think that the answer is yes, and that this is important in understanding *akrasia* in action and intention. Considering *akratic belief* will help set the stage for an account of *akrasia* in action and intention. As Amélie Rorty puts it: "An analysis of the conditions for akratic beliefs brings out some of the hidden conditions of akrasia"(1983, 175).[19] This chapter and the next consider akratic belief, partly to introduce and motivate a way of thinking about *akrasia* in general.

Moreover, the ambitious kind of guise-of-the-good view I am defending *must* consider akratic belief. Recall

The Identity View: A's intention to *x* is a belief that A ought to *x*.

On this view, to have an intention one believes one should not have *is* to have a belief one believes one should not have. *Akrasia* in intention is then itself a kind of *akrasia* in belief. The Identity View cannot account for *akrasia* if, as some think, akratic belief is impossible. If the Identity View is right, we must be able to explain how we can believe against our own 'better' judgment. If the doubts I described above are right, there is no genuine akratic belief; no doxastic analogue to akratic action and intention; and a damaging unsolved problem for the Identity View. If the Identity View is right, akratic belief must be possible.

There is a complication here. It is tempting to try a shortcut, appealing to what is often called "referential opacity." A context or expression is *referentially opaque* when substituting a term referring to the same object as another term may change the truth-value of a statement. For example, Venus is the evening star, but I may not believe it. I can then believe that the evening star is visible, without believing that Venus is visible. The statement "He believes the evening star is visible" is then true of me, while the statement "He believes Venus is visible" is false. In another example, "Lois believes Superman will

[19] Rorty also offers a further motivation for considering akratic belief in the context of akratic action: "Since the best explanation of akrasia of action characteristically lies in akrasia of belief, the best correction for akrasia of action lies in the correction of akrasia of belief"(176). If she is right, an understanding of akratic belief is essential in moral education.

rescue her" can be true, while "Lois believes Clark Kent will rescue her" is false, even though Clark Kent is Superman.

Similarly, on the Identity View, intention is a normative belief. But many people do not believe the Identity View. These people, it seems, can believe that they should not intend to eat dessert, without believing that they should not believe they should eat dessert. In other words, they can have a belief which forbids the intention, and makes it akratic, without having a belief that forbids the first-order normative belief as a belief. The statement "She believes she should not intend to eat dessert" can be true, while the statement "She believes she should not believe she should eat dessert" is false. So even on the Identity View, it seems, there can be *akrasia* of action and intention without *akrasia* of belief. Referential opacity in the 'better judgment' seems to release the Identity View from its commitment to the possibility of akratic belief.

The same thought can be put in terms of the distinction between *de re* and *de dicto* belief attributions.[20] Suppose we see a suspicious-looking man in a trenchcoat. He happens to be the mayor, but we do not know this, and we believe he is a spy. Do we believe that the mayor is a spy? In the *de re* sense, we do: we believe about this man (who is the mayor) that he is a spy. In the *de dicto* sense, we do not: we do not believe that the mayor is a spy, since we do not know that this man is the mayor. That is why, when asked whether we believe the mayor is a spy, we would say no. The same distinction can be made in the case of *akrasia*. Suppose we intend to eat a second dessert, but believe we should not have this intention. If the Identity View is right, this intention is a belief; but we may not believe that the Identity View is right, and so we only see the intention as an intention. (We see, so to speak, a suspicious-looking intention in a trenchcoat.) Do we believe that we should not have this belief? In the *de re* sense, we do. We believe about the intention (which is a belief) that we should not have it. But in the *de dicto* sense, we do not. We do not believe that we should not believe that we should eat the second dessert, because we do not see the intention as a belief. And that, one might think, is the sense that matters. Beliefs about states that we do not even see as beliefs should not count as genuine *akrasia* in belief. It is then too much of a stretch to count the Identity View as committed to the possibility of akratic belief. If that is right, could we not skip these two chapters?

I do not think the Identity View can get away so easily. First of all, there can be people who *do* believe the Identity View. Those people can still act and intend akratically. When they have an intention they believe they should not have, they also believe that the intention is a belief. Unless they *never* put two and two together, they will then believe they should not have that belief. In other words, they will disapprove of that intention/belief state in its belief aspect, not only in its intention aspect. So the Identity View will have at least some akratic beliefs to explain.

There is also a deeper problem. Even the people who do not believe the Identity View would, according to the Identity View, be at least extremely close to *akrasia* in

[20] For influential discussions of the *de re* / *de dicto* distinction, see Quine (1956; the mayor example is his), Kaplan (1968), Burge (1977), Lewis (1979), and Taylor (2002).

belief. Though they might disapprove of an intention of theirs, thinking of it only as intention, they will still have a belief that (under a different guise) they believe they should not have. They will still have *de re* akratic belief. On the Identity View, it is still obscure how this is possible, if 'genuine' or *de dicto* akratic belief is not. If these people believe they should not have the intention, would they not also believe they should not have the corresponding belief? Would they not then at least suspend belief, and in doing so give up the intention? Before accepting the Identity View, we would want to answer tricky questions like these. It is far from clear how *de re* akratic belief is possible, especially if *de dicto* akratic belief is not.

When it comes to avoiding the topic of akratic belief, there is one further problem. If *de dicto* akratic belief is impossible, the Identity View should still provide an understanding of *why* it is impossible, so that we can see whether the impossibility applies to related states like the ones at issue here. That understanding will be important in accounting for akratic action, *de re* akratic belief, and any *akrasia* in people who do believe the Identity View. For all these reasons, the Identity View still has to consider akratic belief, rather than looking for an easy way around it. It does not quite have to say without qualification that akratic intention is a kind of akratic belief. But the upshot is the same: to evaluate the Identity View, we need to ask whether and how akratic belief is possible.

In the rest of this chapter, I consider three arguments against the possibility of akratic belief. For convenience, I call these "impossibility arguments" with respect to akratic belief. The task of this chapter is to show that there is no principled reason to deny that akratic belief is possible. I will argue that all of the impossibility arguments lack independent force, and are instead expressions of an underlying puzzlement about how belief can be akratic. The next chapter will address that underlying puzzlement, drawing on existing views to develop a conception of akratic belief.

II. The Nullification Argument

One natural way to deny that akratic belief is possible—as I believe, the most common way in the existing literature[21]—can be called

The Nullification Argument

(1) To believe that *p*, a person must believe there is reason to believe that *p*.
(2) When someone believes there is conclusive reason to believe *not p*, she cannot believe there is reason to believe that *p*.
(3) So, when someone believes there is conclusive reason to believe *not p*, she cannot believe that *p*.

[21] See especially Hurley (1989), 131-5, Adler (2002a) and (2002b), and Owens (2002). For responses see Scanlon (1998), 35, and Levy (2004), both discussed in this chapter.

The argument can be put in terms of evidence, or in terms of what one believes one should believe, rather than in terms of reasons. The crucial idea in the second premise would stay the same. Once we come to a belief about the reasons to believe *p*, as Adler (2002b, 7) puts it, "contrary or undermining evidence is nullified." Countervailing reasons lose their force, and can have no further hold on us. Owens (2002, 390) similarly writes: "No one can freely and deliberately form the belief that *p* when they think the evidence sufficient to establish its falsehood, because no one can judge that there is *any* reason to believe p in such a situation." Or as Susan Hurley (1989, 131-2) puts it: the contrary evidence "has been subsumed without remainder. Less inclusive probabilistic evidence has no constitutive reason-giving force that could hold out in the face of recognition that it's subsumed by the best probabilistic evidence, which favors the opposite conclusion."[22] I will come back to Hurley's formulation below.

The argument is questionable in several ways. First, (1) can be doubted, and appears to give rise to an infinite regress: to believe that there is reason to believe that *p*, one must believe that there is reason to believe that there is reason to believe that *p*, and so on. Second, (2) assumes that it is impossible to have conflicting beliefs about what we have conclusive reason to believe. If this were not impossible, one could believe there is conclusive reason to believe *not p*, and also believe there is conclusive reason to believe that *p*. The latter would be one way of believing there is reason to believe that *p*. Third, the argument as a whole applies to only one variety of akratic beliefs: those in which the person believes there is conclusive reason for the opposite conclusion. Other akratic beliefs—about God, or abortion, or an upcoming election—might be akratic because we believe we should suspend belief, rather than hold an opposing view. Believing one should *not believe p* does not require belief that one should *believe not p*. Though these difficulties are significant, I will put them aside, in order to, as Wittgenstein put it, attack the opposing view at its strongest point. I will focus on belief in conclusive reason against a belief, because I think we can see *akrasia* even there, in a way that sheds light on the other cases.

Although the nullification argument is the most common impossibility argument in the literature on akratic belief, the crucial second premise is often assumed, and rarely defended systematically. But I think we can piece together what its defenders do say, and understand the idea behind it.

Adler (2002b, 7) supports his view with an example,which we can call

> *The Parking Lot*. If I pass my colleague David's car in the parking lot of the local diner, I conclude that David is inside. But if, within minutes of the observation, I

[22] Levy (2004) borrows Hurley's terminology in calling the view I describe "the subsumption view." As he puts the view: "epistemic akrasia is impossible because when we form a full belief, any apparent evidence against that belief loses its power over us"(149). This, I think, is basically Premise 2, the crucial premise in the nullification argument, and more aptly thought of as a view about nullification rather than 'subsumption'. But talk of subsumption does point to a way in which Hurley's view might offer an argument for Premise 2; see below.

> talk to another colleague on the phone, who mentions that David is in his office, then the evidence of seeing David's car in the lot is nullified. Presuming that I have no reason to distrust the colleague, I infer that it was, e.g., David's wife or son who drove his car to the diner. For if observing David's car retains its original evidential force, then I cannot simply accept it that he is in his office.[23]

For Adler, conclusive evidence (we can imagine seeing David in the office ourselves) nullifies, or cancels, the apparent evidence for the contrary conclusion. If we believe there is conclusive reason to believe David is in his office, we can no longer see his car as evidence that he is in the diner, and so we cannot believe he is in the diner on the basis of that evidence. Any apparent evidence that he is in the diner loses its status as evidence. The car might call for a different explanation, but it no longer supports the explanation that David is in the diner.

Dretske (1971, 216-7) gives another helpful example:

> Bill knows there are some cookies in the jar (he just now looked); Sam does not. Both watch a hungry child peer into the jar, replace the lid without extracting anything, and leave with a disappointed look on his face. Sam now has a reason to think the jar empty; Bill does not. For Bill, who knows there are cookies in the jar, the child's behavior is not a reason to think the jar empty; it is, instead, *something to be explained* (if he does not already know the explanation). Perhaps the child does not like peanut butter cookies. Perhaps he didn't see them when he looked into the jar. What makes the child's behavior something to be explained from Bill's point of view is, of course, the fact that the absence of cookies is not available to explain it. Bill knows there are cookies in the jar, and, hence, knows that the child's empty-handed departure cannot be explained by an absence of cookies.
>
> Bill could take the child's behavior as a reason to think the jar empty…only if he could regard an empty jar as a possible, a more or less competitive, explanation for the child's behavior…. He [then] can no longer be described as knowing that the jar has cookies in it whatever he might have known, or thought he knew, a moment ago. For to persist in saying he (still) knows the jar has cookies in it is to say something absurd: viz. that *S* is treating a hypothesis (no cookies in the jar) which he knows to be false as a possible, a more or less competitive, explanation for the child's behavior.

[23] Adler gives another such example in his book (2002a, 70): "If the initial evidence at the scene of the crime is a scarf, which resembles one that the butler wore, then the evidence, let's say, favors the butler's guilt. But if we subsequently discover that the butler was out of town at the time of the crime, the scarf exercises no further pull on us, even if we cannot discover what it was doing there."

Though Dretske describes his example as involving knowledge, one might think the same holds of belief. If Bill thinks the evidence establishes that there are cookies in the jar, how could he see himself as having any reason to believe the jar is empty? On what basis could he hold that belief? Examples like these make the nullification argument plausible. It is hard to see any reason for believing the jar empty, or David in the diner, when one sees that the evidence against the belief is conclusive.

The nullification argument seems most convincing in cases of *knockdown* evidence—evidence that is both conclusive and especially obvious or impressive. *Seeing* the cookies, or David, or hearing from a reliable person who sees him, does seem to knock down, knock out, discredit, or nullify the apparent evidence to the contrary. But not all examples fit into this narrow range. As Levy (2004) insists, some beliefs are much more complicated. Levy's favorite example is philosophical beliefs. As he puts it: "Dennett embraces compatibilism wholeheartedly;…O'Connor is a libertarian, and so on. Yet, if they are honest, I suspect that they would acknowledge that their position doesn't deal with *all* the cases…in a manner that is completely satisfying…. Some evidence continues to be recalcitrant"(2004, 155). Or as we see in Putnam on his own views: "When I was a 'scientific realist', I felt deeply troubled by the difficulties of scientific realism; having given up scientific realism, I am still tremendously aware of what is appealing about the scientific realist conception of philosophy."[24] We can hold a view while acknowledging the evidence against it. Levy emphasizes that this is not a quirk of philosophical thinking. "Disputes beyond the bounds of philosophy narrowly construed give rise to the same experience. Wherever there is ongoing, rational, controversy,…we shall find precisely this structure of reasons"(156).

I think Levy's objection does not apply to the nullification argument in quite the way he means it to. His examples of philosophical beliefs sound like examples in which the believer does *not* see conclusive reason for his belief. In these examples, Dennett, O'Connor, and Putnam seem to be *holding* beliefs *despite* what they see as a lack of conclusive reason for them—and not, as would be relevant here, seeing conclusive reason but still recognizing opposing reasons. Levy does say that Dennett "wholeheartedly" embraces compatibilism, but this mention is too quick to make its point. We would have to take Levy's word for it, interpret it to mean "with a belief in conclusive reason," and then grant Levy's view that Dennett can still recognize opposing considerations.

Nevertheless, Levy brings out an important point. To see oneself as having "conclusive", "decisive", or even "overwhelming" evidence for a belief is not to think that every single reason or piece of evidence speaks in favor of it. In more complex cases the latter is unlikely even when we have the former. We can see the case for a philosophical view as conclusive if the view is well explained, supported by powerful arguments, and convincingly defended against all major objections, even if we are not yet clear how it handles one of several minor puzzle cases. A jury can convict beyond reasonable doubt despite some evidence of a defendant's good will toward the victim, recognizing that this

[24] Putnam (1988), xii, quoted in Levy (2004), 155.

piece of evidence makes her less likely to have committed the crime. The notion of conclusive evidence or reason does not require eliminating or explaining every last shred of opposing evidence.

Levy's defense nevertheless claims less than I believe it should. Levy insists that there is a broad range of cases in which we can see the force of the evidence against a view we ourselves hold. As he puts it, we can see this whenever there is ongoing, rational controversy. But I think the range is actually much broader than the one he describes. I think we can see the same phenonemenon even where there is only *ir*rational controversy, and even where there is *no* controversy. Seeing this points to a deeper problem for the nullification argument.

Consider the gambler's fallacy. The fallacy in its general form is the belief that if deviations from expected behavior are observed in repeated independent trials of a random process, future deviations in the opposite direction are more likely. For example, if a series of random coin tosses includes several tails in a row, some people start to expect heads. There is no hope of vindicating this prediction beyond a 50% chance; past coin tosses have no effect on future ones! But as we see five, six, seven tails in a row, we can think: it's *got* to be heads next time. In this case, there is no defensible position favoring heads that could be a party in an ongoing rational controversy. The apparent reasons for expecting heads are merely apparent. And yet people who understand the gambler's fallacy can be inclined to take a series of tails as reason to believe the next toss is more likely to come out heads. They can even be inclined to believe it for that reason.

As Descartes and Hume showed us, *mere appearance* can be very powerful. It can lead to belief. If the nullification argument tells us that, as Levy (2004, 152) puts it: "The apparent evidence against *p* is shown to be appearance only, and is therefore stripped of any persuasive force"—then the answer is that appearance can have persuasive force. For all we have seen so far, we may be able to acknowledge this force even when we commit to an opposing view.[25]

Hurley's statement of her view can be taken as an independent argument for (2). She writes (1989, 131-2, quoted above) that the contrary evidence "has been subsumed without remainder. Less inclusive probabilistic evidence has no constitutive reason-giving force that could hold out in the face of recognition that it's subsumed by the best probabilistic evidence, which favors the opposite conclusion." In other words, a parked car

[25] There is a related field of psychological research focusing on the phenomenon of "moral dumbfounding"—"the stubborn and puzzled maintenance of a moral judgment without supporting reasons"(Haidt et al (2000), 1). Most of us, for example, have strong reactions when presented with imagined scenarios of apparently 'innocent' incest or cannibalism (Haidt et al (2000), Haidt (2001) and (2005), Haidt and Bjorklund (2008)). Though it is fascinating and closely related, this research is difficult to use in an argument for the possibility of *akrasia*, since it would be hard to show that the subjects in the experiments believe they should not have the reactions they have. Though the experiments provide striking examples of dissociation between belief and reasoning or justification, they lack consistent reports of contrary normative belief, and so do not clearly show dissociation between belief and belief about what we should believe. For a strikingly incisive critical review of Haidt's experiments, see Jacobson (2012).

or a disappointed look—a set of evidence that includes less—has been "subsumed" by the full set of evidence that includes the rest of what we know. The idea of "less inclusive" evidence being "subsumed without remainder" suggests that the evidence for a view is nullified *because* it has already been taken into account in the judgment in favor of the contrary conclusion. As Levy puts the line of thought, "The evidence against…is exhausted: it is subsumed into our judgment. It therefore retains no further power to move us against our own best judgment. It is absorbed into the set of reasons which support that judgment"(151). According to this view, David's car or the child's disappointment are already taken into account in judgment, and so cannot be brought into consideration in favor of the opposing view.

It is still unclear why not. There are two ambiguities in the description of "subsumption" here. First, is what is "subsumed" the *evidence*—whether thought of as a consideration, fact, or object—or its *force*? Second, is the "force" to be understood as normative, contributing to what we *should* believe, or persuasive, affecting what we *actually* believe? It is tempting to trade on the ambiguities, and say that the evidence loses its *force*, because the *evidence* has already been taken into account; the force can then be understood as persuasive force. Trading on the ambiguities in some such way is essential to Hurley's claim of subsumption. When she says that "less inclusive probabilistic evidence has no constitutive reason-giving force," this is relevant only if taking the evidence into account removes its normative force in favor of what would be the akratic belief, and thereby renders it unpersuasive. But recall Matt's fear of flying, and his question: "What's holding it up there anyway?" Whether a piece of evidence that he believes he should not find persuasive can nevertheless persuade him is at the heart of the issue. And it is not clear why taking a piece of evidence into account in forming one view should rob it of its persuasive force in supporting another. We can take into account a car, or a look, or a set of statistics about airplane safety, and these might, for all we have seen, play the role of opposing evidence as well. There is so far still no reason why evidence cannot play these two roles. Most importantly, we have not seen why persuasive force should match what one believes it should be.

To say this is not yet to explain *akrasia*. But it does bring out the possibility that we can believe for weaker or merely apparent reasons, even when we believe in conclusive reasons for an opposing view. The nullification argument appears convincing only because we cannot yet see how we can believe something against our own view of what we should believe. It seems convincing, in other words, because *akrasia* is so puzzling. The argument then has no independent normative or persuasive force. It is a dramatic but derivative expression of the difficulty of understanding akratic belief.

III. The Argument from Moore's Paradox

The early twentieth century British philosopher G.E. Moore noticed the oddity of statements like: "It's raining, but I don't believe it." Such an assertion strikes most people

as odd, absurd, incoherent, or even nonsense. On hearing it, we might ask what the speaker means. It is natural to doubt that she could be sincere.

The absurdity is itself puzzling. Suppose it is raining, but our friend in another room does not believe it. This is entirely possible, and there is nothing contradictory about supposing it. She might have seen a misleading weather report, believed someone's deception, mistaken an artificial light for the light of the sun, or just not have looked outside yet. It can be true, at the same time, both that it is raining and that she does not believe it. We can intelligibly say this about her; "It's raining, but she doesn't believe it" does not strike anyone as absurd. Nor does a statement about one's own past; she can say later "It was raining, but I did not believe it." But there is something strange about her saying it about the present—even when it is true. As Moran (2001, 69-70) puts it, ignorance of one fact or another, and error in one or more beliefs, are not states we can reasonably hope to outgrow; and yet we cannot straightforwardly report them in particular cases. Moore (1993, 209) presents the paradox in these two interrelated ways: both as a paradox about why the problem does not arise in the past or third-person, and as a paradox about why "it should be perfectly absurd to utter assertively words of which the meaning is something which may quite well be true – is not a contradiction." According to Moore, it is *absurd* to assert "It's raining, but I don't believe it"; the *paradox* is that such an assertion is absurd only in the first-person present tense, and despite the fact that what it asserts is consistent and can be true.[26]

Despite its limitation to the first-person present tense, Moorean 'absurdity' comes in different forms. It arises in "It's raining, but I don't believe it," a report of a lack of belief, and also in "It's raining, but I believe it isn't." These two versions are often called *omissive* and *commissive*; a statement of the form "P, but I do not believe P" reports omission or ignorance, while "P, but I believe not-P" reports committing a mistake. We can also find other variants concerning knowledge, certainty, or even guessing— "It's raining, but I don't know it / but I know it isn't"; "It's raining, but I'm not sure it is / I'm sure it isn't"; "It's raining, but I don't guess it is / I'm guessing it isn't"—that are odd or absurd in at least related ways. And the paradox applies to both assertion and belief. Just as

[26] Moore's paradox first arises in Moore (1944a, 175; 1944b, 543; and 1993). In the first two of these, Moore mentions it mostly in passing, as a way to illustrate a distinction between asserting and implying. In the last, an unpublished talk dated by its editor to 1944, Moore gives both presentations of the paradox, of which he thinks the second, quoted here, is "the fundamental one"(209). Sorensen (2007, 47-8) traces the paradox, and the distinction between asserting and implying that it was meant to illustrate, to Moore's resistance to analysis of concepts in purely subjective terms, and so to his rejection of idealism and expressivism.

Malcolm (1958, 66) reports Wittgenstein remarking that "the only work of Moore's that greatly impressed him was his discovery of the peculiar kind of nonsense involved in such a sentence as, e.g., 'It is raining but I don't believe it.'" Wittgenstein devoted (1953), Part II, §x and parts of (1980a; esp.§§460-504) and (1980b; esp. §§278-90) to the paradox, drawing out its importance, giving it the name "Moore's paradox"(1953, 190-1), and bringing it much wider attention.

it is odd to say: "It's raining, but I don't believe it," it is hard to imagine someone *believing* such a thing. This version of the paradox will be especially important here.[27]

It is common to mention Moore's paradox and akratic belief in the same breath. Akratic belief can be seen as involving a kind of Moorean absurdity. Huemer (2007, 146) includes as a Moore-paradoxical statement: "It is raining, but I have no justification for thinking so." Gallois (2007, 166-7) writes: "Additional examples of Moore-paradoxicality are provided by… 'P, but I am not at all justified in believing that P.'" According to De Almeida (2007, 56), "P, but it's not rational for me to believe that p" and "P, but it's rational for me to believe that not-p" are a kind of "doing what, by your own lights, you shouldn't be doing," and "intuitively seem to be no less cases of Moorean absurdity" than the standard cases. For Adler and Armour-Garb (2007, 161-2), "Instances of the following are variants of M[oore's] P[aradox]…: *p*, but I lack sufficient evidence that *p*. *p*, but my reasons do not establish *p*…. Any statement of the form '*p* but I M that *p*,' if all-out believed, will be a version of Moore's Paradox, if M serves to cancel the grounds or reasons for fully believing that *p*." This general schema presumably includes "*p* but I should not believe that *p*" as a central instance of cancelling the grounds for fully believing that *p*.

By itself, a brief mention of Moore's paradox in a discussion of akratic belief shows little. As I will explain, I think it is not a coincidence that these mentions have been brief. But it is striking that the association is so natural, and that it arises so naturally in the context of a denial of the possibility of akratic belief. These denials can include a gesture at Moore's paradox in a central, if not yet fully explained, role:

> Imagine that your beliefs run counter to what evidence and fact require. In such a case, your beliefs will not allow those requirements to remain visible because the offending beliefs themselves give you your sense of what is and your sense of what appears to be. You are therefore denied an experience whose content is that you are believing such-and-such in defiance of the requirements of fact and evidence. This is why, as G.E.Moore observed, you cannot simultaneously think that while you believe that *p*, yet it is not the case that *p*.

---

[27] Moore introduced both the omissive "I went to the pictures last Tuesday but I don't believe that I did"(1944b, 543) and the commissive "I believe he has gone out, but he has not"(1944a, 175), as well as the now standard rain example in the omissive form "I don't believe it's raining, but as a matter of fact it is"(1993, 207). But neither Moore nor Wittgenstein explicitly distinguished or contrasted the omissive and commissive forms of the paradox, and other early treatments of it often applied to one but not the other.

Moore and Wittgenstein also limited their treatment to assertion rather than belief. Sorensen (1988), Shoemaker (1995), and others have come to see the paradox about belief as more fundamental. Variants concerning higher-order beliefs and other mental states also abound in the literature, and it is even a bit loose to say, as I do, that the paradox only concerns the present tense; for a discussion of the future tense variant, see Bovens (1995). Explanations of the paradox come in even more varieties; Green and Williams' introductory discussion of 18 different types in their (2007) is a useful starting point.

Pettit and Smith (1996), 448

> (2) p, even though the evidence indicates not-p
> …[is] heard as suffering a Moore's Paradox-type contradiction that is explained as due to a corresponding incoherence in thought. If assertions of the form of (2) are Moore's Paradoxical, then…it would be surprising if any model of akratic action that retains first-personal irrationality can serve as a model of akratic belief. For the first-personal thought corresponding to the admission of akratic belief would be not merely irrational, but incoherent.
>
> Adler (2002a), 21

The first, dense passage from Pettit and Smith does not say why our beliefs about what we ought to believe must give us our "sense of what is."[28] Adler, who concludes his denial of the possibility of akratic belief with the second passage, does not say why his (2) is "heard as suffering a Moore's paradox-type contradiction," or why it would be right to hear it this

[28]Owens (2002), 382-3 includes a brief rejection of impossibility arguments from Moore's Paradox, taking the passage from Pettit and Smith as his main example. He criticizes them for conflating fact and evidence (382):

> (1) I believe Jones is innocent but this belief is based on insufficient evidence
> …is not equivalent to
> (2) I believe Jones is innocent but he is guilty.
> …Pettit and Smith appear to equate (2) with (1) and thence infer that the state of mind expressed by (1) must be impossible also. But if the state of mind expressed by (1) is impossible, it is not for this reason.

Though Owens makes this criticism with consideration of both commissive and omissive assertions, his criticism is softened both by its focus on the distinction between fact and evidence, and by its legal example. Talk of legal evidence such as an "alibi" and "eyewitness testimony"(383) can be misleading, since it can prompt us to imagine, for example, Jones' character, as we know it outside court, to be the operative contrary evidence in the epistemic sense. The distinction between kinds of evidence is itself dangerous, to the extent that it cements the assumption that the notion of evidence is the right central notion. I think it is not. In cases of religious belief, for example, someone can believe that God exists, and believe that she should believe it, even though the evidence is insufficient. That believer displays not *akrasia* but a particular view of the relation between evidence and norms for belief. Taking 'ought' or 'should' (I continue to treat these as synonymous) as the central notion instead of 'evidence' avoids blurring these delicate issues.

Most importantly for the analogy with Moore's paradox, I doubt Pettit and Smith can be convinced so easily. Even without a clear view of the distinction between fact and evidence, they seem to think that the requirements of evidence cannot "remain visible" in the presence of a conflicting belief any more than the requirements of fact can. More generally, there remains a sense that it is somehow Moore-paradoxical to go against what we believe is required of us. This sense is expressed in the other passages I quote in the text, which, though not themselves systematically defended, do not face Owens' criticism of Pettit and Smith. In the text I avoid criticizing their quick and distracting formulation of the argument from Moore's paradox, and focus instead on reconstructing the argument in its strongest form before explaining why I believe it cannot succeed.

way. But both passages gesture at Moore's paradox, with a sense of the importance of the connection.

The connection has never been defended systematically. But since the association is natural and often made, it is worth asking whether Moore's paradox can provide an argument against the possibility of akratic belief. I think it is not accidental that these gestures have been so brief and unsystematic. In the rest of this section, I will develop the parallel in more detail, and argue that there is no independent argument here against the possibility of *akratic* belief. In other words, I will develop the natural appeal to Moore's paradox, in order to show why it cannot be successful.

To develop the analogy to Moore's paradox, we can look for a formulation of *The Belief Akratic's Paradox*. One example of a belief-akratic-paradoxical assertion might be: "I believe it's raining, but I shouldn't believe it." As in Moore's paradox, the absurdity depends on the first person present tense; "She believes it's raining, but she shouldn't believe it" is not odd or absurd, and neither is: "I believed it was raining, but I shouldn't have." Like Moore's paradox, this one also allows variations involving other states: "I'm certain it's raining, but I shouldn't be"; "I'm guessing it's raining, but I shouldn't"; these are variations on the paradoxical theme, though not quite the standard case. And as in Moore's paradox, it would be too quick to assume that there is only one standard formulation. Since we are evaluating an analogy that has never been systematically developed, it is worth stopping to ask what the basic forms of the paradox will be.[29]

An akratic belief is a belief that one believes one should not have. The paradoxical statement should express or report both of those beliefs: the akratic belief itself, and the belief that forbids that belief and renders it akratic. Whether we treat *expressing* or *reporting* as central already makes a difference. Simply saying "It's raining" can express a belief that it is raining. But this expression does not report the belief; only "I believe it's raining" does that. An expression of both beliefs might be: "It's raining, but I shouldn't believe it." A report of both would be: "I believe it's raining, but I believe I shouldn't believe it." It may seem obvious that we can at least ignore hybrids of expression and report, such as: "It's raining, but I believe I shouldn't believe it." But my first example—"I believe it's raining, but I shouldn't believe it"—is itself such a hybrid, reporting the akratic belief but merely expressing the prohibitive one.[30]

Does the Belief Akratic's Paradox have omissive and commissive versions? A denial of the first belief would yield: "I do not believe it's raining"—which, to preserve the paradox, would be followed by: "but I should believe it." We would then have: "I do not believe it's raining, but I should believe it," or, less awkwardly: "I should believe it's

[29] I call these assertions and beliefs "belief-akratic-paradoxical" to bring out the parallel with the assertions and beliefs with which Moore's paradox is concerned. But since it is often forgotten even in Moore's case, it is worth noting again that, strictly speaking, the assertions and beliefs are not themselves paradoxical. They are, as Moore put it, 'absurd'; the paradox concerns the limitations on the absurdity, which I considered earlier in this section.

[30] For an excellent discussion of expressing and reporting, see Moran (2001), esp.100-107. My use of "mere" follows Moran in avoiding the view that a report cannot also be a kind of expression.

raining, but I don't." What this statement suggests about its speaker is puzzling, and may count as akratic, but it is not akratic belief. It is akratic suspension or lack of belief: an akratic *not* believing, while believing one should in fact have a belief. This is an interestingly related paradox, but not a formulation of the Belief Akratic's Paradox, since it involves no akratic belief at all.[31] We can try instead a denial of the second belief: "I believe it's raining, but I don't believe I should believe it," or: "It's raining, but I don't believe I should believe it." This does sound odd, and we might want to ask the speaker why she thinks it's raining, if she doesn't think she should believe it. But it is once again not akratic belief. The speaker does not display a belief she believes she should not have. She displays a belief that she may not have a further evaluation of at all. Its analogue in action is then not *akrasia* but motivation without evaluation (see Chapter 5). There is, then, no way to insert omissions, or denials of belief, into the Belief Akratic's Paradox, while preserving the *akrasia* in belief that is central to it. This paradox thus allows a different set of variations than Moore's. It has expressive and reportive versions, but no omissive one. All of its forms must be commissive.[32]

The basic forms of the Belief Akratic's Paradox are then as follows:

Expressive: "It's raining, but I shouldn't believe it."
Reportive: "I believe it's raining, but I believe I shouldn't believe it."
Hybrid 1—expressive-reportive: "It's raining, but I believe I shouldn't believe it."
Hybrid 2—reportive-expressive: "I believe it's raining, but I shouldn't believe it."

---

[31] Owens (2002, 383) proposes "a broader notion of 'epistemic akrasia'," which includes failing to believe. I think he and I would agree that akratic suspension of belief does not count as akratic believing, but does count as *akrasia* pertaining to belief (though 'doxastic' would be a more appropriate label than 'epistemic', or pertaining to knowledge), and so is worth keeping in mind. The difference is mainly in emphasis. I nevertheless focus on akratic belief as the central case, partly because, if we conclude from the possibility of akratic suspension of belief that "epistemic akrasia is possible," this can distract us from the question of whether akratic belief is possible.

Because I do not consider suspension of belief in the text, I also do not consider the distinction between suspending and merely lacking a belief, the way most of us simply lack beliefs about the size of some nearby stars. But I have this distinction in mind when I mention "suspension or lack," and say that it "may" be akratic. Mere lack of a belief one believes one ought to have may not properly be called akratic; but I do not take a stand on that here.

[32] I do not mean that the omissive versions are not paradoxical—only that they are not akratic. They might be called the Non-Believer's Paradox and the Non-Endorsing Believer's Paradox, respectively. These continue the paradoxical theme, without involving beliefs one believes one should not have.

There is one other variant which can sound omissive: "I believe it's raining, but I believe it's not true that I ought to believe it." This version differs from the reportive one by denying an 'ought' instead of asserting an 'ought not', allowing that the belief might be neither required nor prohibited but merely permitted. This is distinctive, but not an omission, since it reports a particular belief, rather than the absence of one. Nor is it akratic, since it involves no belief that the believer believes she should not have.

All of these sound odd, and all are at least closely connected to akratic belief. All are odd both as assertions and as beliefs, though the corresponding beliefs do not themselves involve outward 'reports'. And in each case, one can see why philosophers would think of Moore. Each variant is a distinctively first-person, present-tense conjunctive statement or belief that is not self-contradictory and yet seems to somehow involve or implicate the speaker or believer in contradiction, absurdity, or nonsense. The Belief Akratic's Paradox, one might think, is like Moore's but worse. It does not just make a commitment and undermine it in the same breath. It expresses (and in some cases, reports) two beliefs, one of which explicitly forbids the other.

We can use the term "Moorean belief" for a belief expressed by an assertion that is odd in the way Moore drew attention to: e.g., the belief that (it's raining, but I don't believe it). Similarly, we can call beliefs expressed by an assertion in the Belief Akratic's Paradox "belief-akratical-paradoxical beliefs." An argument appealing to the analogy can then be reconstructed as follows:

The Argument from Moore's Paradox

(1) Moorean belief is impossible.
(2) Akratic belief requires belief-akratic-paradoxical belief.
(3) If Moorean belief is impossible, then belief-akratic-paradoxical belief is impossible.
(4) Belief-akratic-paradoxical belief is impossible. (From (1) and (3).)
(5) If belief-akratic-paradoxical belief is impossible, then akratic belief is impossible. (From (2).)
(6) So, akratic belief is impossible. (From (4) and (5).)

This argument is valid. (4)-(6) are easy inferences from premises (1)-(3), and they reach the needed conclusion.

I will concede (1), the impossibility of Moorean belief. I believe Moorean belief is possible, and one kind of example of it will emerge in the next section. But I will not rest the answer to the Argument from Moore's Paradox on the possibility of Moorean belief. This makes things more difficult for a defender of the possibility of akratic belief, but it also keeps the focus on *akrasia*, avoiding a systematic consideration of the literature on Moore's paradox that would itself be a book-length project. Still, it is worth noticing that the impossibility of Moorean belief is essential for the argument, and that this itself is controversial.[33] Some of us think that Moorean beliefs are themselves odd, irrational,

[33] Though he did not consider Moorean beliefs, Moore said of Moorean assertion that "It's perfectly absurd or nonsensical to say such things"(1993, 207). Although 'absurd' is itself ambiguous between 'ridiculous' (but possible) and 'unintelligible' (and therefore impossible), Moore seems to have in mind the latter. Wittgenstein similarly wrote in a letter to Moore that "It makes *no* sense to assert 'p is the case and I don't believe p is the case'"(see Baldwin (1990, 226-32) for discussion). So both can be read as likely to endorse (1), since the corresponding assertion

unstable, incoherent, absurd, or some combination of these, but not impossible. In that case the impossibility argument from Moore's paradox could not even begin. But I will try to answer the argument in a different and shorter way, by arguing that, even if Moorean belief is impossible, the analogy with *akrasia* does not hold. I will argue that (2) is false, and (3) is question-begging.

(2) is, I think, the assumption that remains conveniently hidden when one merely gestures at Moore's paradox in denying the possibility of *akrasia*. It might seem obviously true. Doesn't the paradox express exactly the akratic state? But in fact, akratic belief is quite different from belief-akratic-paradoxical belief. An akratic belief is a belief that one believes one should not have. *Akrasia* in belief is the state of having a belief one believes one should not have—i.e., both some particular belief, and a second belief that one should not have the first one. In any of its four versions, a belief-akratic-paradoxical belief would be a *third* belief, distinct from the other two.

To take just one example, consider the reportive version: "I believe it's raining, but I believe I shouldn't believe it." To believe this is to believe that one has both component beliefs in the akratic state. One must make a substantive assumption to think that, if one has an akratic belief, along with the second belief forbidding it, one must have the third belief *that one has the first two*. Akratic belief does not require that further, third belief. In general, having two beliefs does not require having a third one about those two.[34]

---

seems to have nothing to express. But there is the possibility of seeing Moorean beliefs as possible but inexpressible, and, more importantly, Wittgenstein's later qualification that under some circumstances Moorean assertions can have a sense.

For more explicit endorsements of (1), see Hintikka (1962, 67); Van Fraassen (1984, 247): Shoemaker (1996, 85-6); and Goldstein (2000, 86). Shoemaker writes that "Belief in such a proposition is impossible"(1995, 222), but then, prompted by Albritton (1995), retreats to saying that these propositions "could not be *coherently* believed"(1995, 227n1). Acceptance of the possibility of Moorean belief is nevertheless widespread, including de Almeida (2007), Kriegel (2004), Sorensen (1988), and others, whose discussions of the paradox usually aim to explain not impossibility but irrationality. More recent arguments that akratic belief must be irrational, such as those in Greco (2014) and Horowitz (2014), also tend to assume its possibility.

[34] Such a general "third belief" requirement would have two unacceptable consequences. First, it would generate an infinite regress. The third belief, together with the second, would also be two beliefs. To have those, one would then need a fourth belief about the second and third. That would require a fifth belief, and so on *ad infinitum*. Every randomly selected pair of one person's beliefs would have an infinite number of beliefs attached to it. If that idea is coherent at all, it is unlikely to be true.

This first, formal problem points to a second one. The general requirement ignores the psychological separateness of a person's beliefs. Suppose our friend in the other room believes that it is raining, and that Mongolians use a Cyrillic alphabet. These two beliefs may never have been considered at the same time, or interacted in any way. She may never have "put two and two together" to form a belief about those two beliefs. There is so far no reason to attribute to her a third belief about those two. There may sometimes be such a reason, once a connection is seen. If I am late and running for the bus, but recall that the train is faster, I can more easily be said to believe that I believe both that I am late, and that the train is faster than the bus. But to think that, for any two beliefs of mine, I must have a third belief that I have those two, is to treat all beliefs as connected, if not immediately in awareness, then at least by further belief. There is little motivation

One might object that akratic beliefs are a special case, in which such a requirement does hold. After all, does akratic belief not require a second belief that one should not have *that first one*? The two beliefs in the akratic state are much more closely connected than most pairs of beliefs. In "I should not believe it," the 'it' already refers to the first belief, that it is raining. If one of the beliefs is about the other, it can be hard to see how one can not "put two and two together" to form the third belief. Does this not show that *akrasia* does put us in a belief-akratic-paradoxical state?

This objection rests on at least two mistakes. The first mistake is ignoring the distinction between believing that one should not believe something, and having the further, third-order belief that one has that belief. These are still distinct. One is a belief about a belief; the other is a belief about a belief about a belief. There is still no reason to think that, whenever we believe we should not be in some state, we also believe that we believe we should not be in that state. Even in this somewhat less general form, the assumption is unsupported and unlikely. The second mistake is assuming that either the further belief or the original prohibitive one must give rise to a further, conjunctive belief. This belief too is different from the prohibitive belief itself.[35]

---

for such a general view. But, in the reportive case, that is what premise (2) must assume about akratic beliefs.

[35] This mistake remains even if we can justify the first 'mistake' by showing that belief is essentially self-aware, and either gives rise to or is partly constituted by a belief that one has the belief. Such self-awareness would not require awareness of conjunctions of which the belief provides only one component.

A third mistake is failing to distinguish the content of a belief from the state itself—or, in other words, what is believ*ed*, from the believ*ing*. This distinction is left out in the simple description of akratic belief as a belief one believes one should not have. But it is important here. Consider these two pairs of beliefs:

It is raining; I should not have this belief of mine.
It is raining; I should not believe that it is raining.

The second of these is enough for *akrasia*, without assuming that I do believe it is raining. Someone who has both beliefs in the second pair believes that is raining, but believes that she should not believe it. The second belief is about the rain, and about believing that it is raining. But it does not require believing or noticing that one has the first belief oneself. In the less glaring cases of *akrasia*, one can have both components of the akratic state without yet "putting two and two together" in the sense of believing that one has both. This is how we tend to see akratic action as well. I can deliberate, from my belief that I should not eat sweets, to the conclusion that I should not eat this very sweet in front me. This conclusion is about what is done, not about my doing of it; it is not the conclusion that this action of mine is something I should not be doing. It is simply the conclusion that I should not eat this. But we tend to think that this is enough for *akrasia*. When I believe I should not eat this sweet, and I eat it anyway, I am being akratic.

There is another way to see this point. We tend to think that the prohibitive belief survives even if we give up the akratic one, and can be present before we have it. If I give up the belief that it is raining, in response to my belief that I should not believe it, I go on believing that I should not believe it. My remaining belief is not only about the past; but the second belief in the first pair

The point about the formation of a conjunctive belief applies to all four versions of the Belief Akratic's Paradox. All of them are conjunctive; and in each case, akratic belief does not essentially require a conjunctive belief. So far, this leaves open the possibility that the most extreme cases of akratic belief do involve belief-akratic-paradoxical belief. What does require the third, belief-akratic-paradoxical belief is what we might call "clear-eyed" *akrasia* in belief—the kind in which one sees clearly that one has both component beliefs, and nevertheless maintains both. I will soon turn to the possibility of such cases. But akratic belief in general does not require a third, belief-akratic-paradoxical belief. (2) is false.

At this point one might want to rephrase (2) to rely only on the *possibility* of belief-akratic-paradoxical belief. (2) could then become

> (2*) If akratic belief is impossible, then belief-akratic-paradoxical belief is impossible.

This revision would preserve the validity of the Argument from Moore's Paradox. Moreover, it can be thought to have a deep and general motivation: the thought that, if someone is in a mental state, it must be possible for her to believe she is in that state. It would be odd if we could have akratic beliefs, without being able to believe we have them.

This revised Argument from Moore's Paradox would still depend on the impossibility of belief-akratic-paradoxical beliefs, which I will come to with premise (3). But even if belief-akratic-paradoxical beliefs are *not* possible, there is still good reason to reject (2*). Suppose that belief-akratic-paradoxical beliefs are not possible. Akratic beliefs, if there are any, would then depend on the believer's ignorance of having both the akratic belief and the prohibitive one. She could not "put two and two together" to form the belief that (she believes that her plane will crash, but she shouldn't believe it). Akratic belief would cease to exist as we become aware of it. If *akrasia* depends on ignorance, this is not surprising. States like ignorance or forgetting are paradigm cases of states that cease to exist as we become aware of them. These states undercut the general motivation for (2*); there do seem to be states we cannot believe we are in while we are in them. Most importantly, the impossibility of belief-akratic-paradoxical beliefs would not show that such essentially unaware *akrasia* is not possible. (2*) is then at best a question-begging premise, which assumes in advance that unaware akratic belief is not possible. And if akratic belief is possible—a possibility (2*) cannot help undermine—and belief-akratic-paradoxical belief is not, then (2*) is false. In either case, whether or not belief-akratic-paradoxical belief is possible, (2*) cannot help the Argument from Moore's Paradox.

We can turn to (3): "If Moorean belief is impossible, then belief-akratic-paradoxical belief is impossible." This conditional statement is of course true if Moorean

---

above is, since it has no current belief left to refer to. This suggests that the prohibitive belief is from the second pair.

beliefs are *not* impossible. But the argument needs it to be true even assuming that Moorean beliefs are impossible.

It may seem unfair to think (3) begs the question by simply adding the consequent. The thought (3) gestures at is that beliefs expressed in the Belief Akratic's Paradox are impossible *in the same way* as Moorean beliefs. In other words, they share a feature which, if impossible in one case, would also be impossible in the other. The thought is that the impossibility of Moorean beliefs carries over to the impossibility of akratic ones. This is an appealing thought. But it is worth noticing that, despite its similarities, the Belief Akratic's Paradox is quite different from Moore's.

The belief: 'It is raining, but I believe it isn't' requires a belief that it is raining and a belief that one believes that it is not. (Here I assume that belief distributes over conjuncts.) The belief: 'It is raining, but I don't believe it' requires a belief that it is raining and a belief that one does not have that first belief. Whatever the explanation of the paradox, it is at least natural to think that a contradiction is at the heart of it.[36] The Moorean believer seems to be believing and denying the same thing at the same time.

The belief-akratic-paradoxical believer is in a different position. Recall the four forms of the belief:

> *Expressive*: "It's raining, but I shouldn't believe it."
> *Reportive*: "I believe it's raining, but I believe I shouldn't believe it."
> *Hybrid 1—expressive-reportive*: "It's raining, but I believe I shouldn't believe it."
> *Hybrid 2—reportive-expressive*: "I believe it's raining, but I shouldn't believe it."

None of these reject or deny having a belief that they also express or report. Nor could they in general depend on an underlying contradiction. If it were true that the only way to believe that one should not have a belief is to believe that it is false, this would bring the Belief Akratic's Paradox closer to Moore's. But, as we saw in the previous section, this assumption is false. Many akratic beliefs—about God, or abortion, or an upcoming election—are akratic because we *believe we should suspend judgment*, rather than hold an opposing view. We may simply believe we do not have access to the relevant evidence, either in principle or temporarily.

Even if contradiction is at the heart of Moore's paradox, it cannot be at the heart of the Belief Akratic's. So if belief-akratic-paradoxical beliefs are impossible, they are not, as (3) suggests, impossible in the same way as Moorean ones are. Despite its structural similarities, the Belief Akratic's Paradox concerns beliefs in two distinct propositions, neither of which is a simple negation of or denial of believing the other. This leaves open

[36] I say "at least natural" because I have not systematically considered alternative explanations of the impossibility or irrationality of Moorean belief. But it is striking how many of the existing explanations (usefully surveyed in the Introduction to Green and Mitchell (2007)) appeal to some kind of underlying contradiction. I leave the details out here, because the disanalogy cuts across them. Belief-akratic-paradoxical belief does not require denying either a believed proposition, or that one believes it.

even the possibility of clear-eyed *akrasia* in belief, in which the believer is fully aware of having both component beliefs. We have not yet seen how any of this is possible; but we also have not seen why it would not be.

Akratic belief does not require *any* version of the beliefs or assertions in the Belief Akratic's Paradox. And this paradox itself differs in important ways from Moore's, making it unclear why, if Moorean beliefs are impossible, the impossibility should carry over to akratic ones. After developing the analogy between the paradoxes, we still have no independent argument against the possibility of akratic belief. What we have is another expression of underlying puzzlement. We do not yet understand how belief can be akratic.

IV. The Argument from Transparency

There is another way to defend the sense of underlying similarity between akratic belief and Moore's Paradox. Beliefs, or reasoning or questions about them, are often described as *transparent*, in the sense that each of us comes to a belief about whether she believes that *p* by coming to a belief about whether *p*. One so to speak "looks through" questions about our own psychology and out to what our beliefs are about. When asked: "Do you believe it's raining?", you think about the weather. As Moran (2001, 62) puts it: "A first-person present-tense question about one's belief is answered by reference to (or consideration of) the same reasons that would justify an answer to the corresponding question about the world."[37]

Transparency holds only in the first-person present. We cannot answer questions about others' or our own own earlier beliefs about rain simply by thinking about whether it is or was raining at the time. Transparency seems to hold, in other words, in the same range of beliefs that are subject to Moorean absurdity. And it can seem to explain the absurdity. I settle the question whether I believe it is raining by settling the question whether it is raining; so I seem to be settling it both affirmatively and negatively (or, in the omissive case, at least agnostically) when I believe or say "It's raining but I don't believe it," or "It's raining but I believe it isn't."

[37] Moran traces the term 'transparency' to Edgley (1969), 90. For references to other literature see Byrne and Boyle (2011, esp. 23), who trace the term back to G.E. Moore's well known discussion of the 'transparency' of sensation. There is some variation in the literature in descriptions of what transparency is. Moran (2001, 61) criticizes Edgley's characterization of transparency in terms of our inability to distinguish two questions from each other. Adler and Armour-Garb (2007) characterize transparency in terms of what's true, from one's "point of view"; I avoid this, since it is hard to make sense of the notion of what's true from one's point of view except in terms either of a relativistic notion of truth, which they do not have in mind, or of what one believes, which would make transparency trivial. As I describe it, a belief, and a question about belief, can both be described as transparent if a question about the belief is answered by answering a different question. My usage is stipulative; it allows both beliefs and questions about beliefs to be called "transparent," and I do not claim that everyone uses the term this way. Those who prefer to limit the bearers of transparency either to beliefs, or to questions about them, but not both, can adapt what I say to fit an alternative usage.

The same feature can seem to provide an independent explanation of the impossibility of akratic belief. One might think that both the question of what I believe, and the question of what I should believe, are transparent to the corresponding question about the world.[38] In that case, we answer the normative and psychological questions in the same way. When I ask: "Do I believe it's raining?" or "Should I believe it's raining?", I look to the evidence of rain. There is then no way to answer one question affirmatively and the other negatively.

This line of thought can seem to offer some hope for reviving the Argument from Moore's Paradox. If it is right, akratic and Moorean beliefs may be impossible in the same way. Both would lack transparency, a key feature of belief in general. This would provide independent support for premise (3) of the argument in the previous section.

I will not go into detail about the relation between transparency and Moore's paradox, or insist that transparency provides the key explanation. The crucial point here is that transparency can seem to provide an independent explanation of the impossibility of akratic belief. If it does, and Moore's Paradox otherwise does not, then Moore's Paradox would merely offer a suggestive analogy, while transparency would be at the heart of the impossibility argument. That is why I treat this as a distinct line of argument, and give it a separate treatment here.

What would the explanation be? Owens (2002, 384) suggests one possibility:

> How might one move from the transparency of belief to the impossibility of thinking your own beliefs to be unreasonable? One line of thought goes as follows: the way to form a belief on a given topic is to work out what the truth is, and the way to do that is to look for evidence sufficient to establish the truth. But exactly the same method is used to work out whether a given belief would be reasonable. So the method you use to determine what is the case must deliver the same result as the method you use to discern what it would be reasonable for you to believe. How then can you end up with a belief which you yourself think to be unreasonable?

According to this line of thought, though the question about truth is not answered *by* answering the question of what is 'reasonable' to believe—i.e., is not itself transparent to a normative or evaluative question—the two are answered in the same way. Both look to the evidence for the belief. So if the question of whether I believe something is transparent to the question of whether it is true, the answer cannot diverge from the answer to the normative or evaluative question. The two answers are reached by a single procedure, which stays the same in both cases.

We can call this

[38] See, e.g., Shah and Velleman (2005, 502), and for discussion, Owens (2002, 384).

The Argument from Transparency

(1) A person comes to a belief about whether she believes that *p* by coming to a belief about whether *p*.
(2) A person comes to a belief about whether *p* by considering the evidence for *p*.
(3) A person comes to a belief about whether she ought to believe that *p* by considering the evidence for *p*.
(4) So, a person's beliefs about whether she believes that *p*, whether *p*, and whether she ought to believe that *p* cannot diverge.
(5) So, it is impossible to believe that *p* and believe that one ought not believe that *p*.

Here the starting point, in premise (1), is transparency, and the line of thought is the one described by Owens. To reach the conclusion, "answering a question" must be understood as coming to have a belief, rather than as an outward act of assertion. For brevity, I use the term "coming to a belief," though this can include both initial formation of a belief and reaffirming a belief one already has.

Like the Nullification Argument, the Argument from Transparency is questionable in several ways. I will mention three only to set them aside. First, the argument has a hidden premise ruling out the possibility of conflicting beliefs about what we ought to believe. If someone could believe she ought not believe that p, and *also* believe she ought to believe that *p*, she could still come to believe that *p* in coming to that second normative belief. I will put much more emphasis on the possibility of such conflict in Chapter 4, but it is worth noticing the way this argument ignores it. Second, to reach (4) and (5), "comes to a belief" must be meant in the broad sense that includes all belief formation, not just the formation of beliefs by deliberation. But then (2) is hard to accept in its current form, since not all beliefs are formed through a *consideration* of evidence. No reflection need occur. Third, some beliefs formed through self-deception or wishful thinking may not be based on evidence at all. An added requirement, that every such belief still be accompanied by a belief that one ought to have it, can seem to be unmotivated, and to give rise to an infinite regress—one would also have to believe that one one ought to believe that one ought to believe it, and so on. I leave aside the question of how these doubts can be addressed, to focus on the underlying motivation for the argument.

Owens himself finds the line of thought he describes unconvincing. To resist it, he introduces examples of emotion to bring out the way in which an attitude and a belief that the attitude is 'unreasonable' can arise in response to the same evidence. Anger, for example, can be a reaction to evidence of a friend's incompetence in fixing my computer. But if I believe the evidence does not give me good reason for my anger, I can believe the anger is unjustified, in response to the same "ineffectual tappings, etc." that infuriate me (385). Owens also thinks "it is worth pointing out that the argument from transparency would make practical akrasia seem no less problematic than theoretical akrasia"(386). Like

anger, desire and intention can arise in response to the same features of the world as the belief that makes a decision akratic.

Owens' response is entirely concerned with attitudes other than belief. The focus on other attitudes leaves open the possibility that belief is simply unlike them—that anger, for example, is not transparent in the same way as belief. A defender of the possibility of akratic action and intention could similarly insist that they are not transparent. It would not be hard to make the case for a disanalogy. In the case of these other attitudes, it is hard to see how questions about them could *ever* be transparent. The question of whether I have some emotion or intention is hardly settled *only* by looking at the evidence in the way that we do to determine whether the object of the emotion or intention is true. The question is certainly not transparent to the corresponding question about the world. If asked whether I am angry that it is raining, I cannot answer this question simply by answering the question whether it is raining. So it seems that with anger, it cannot be enough to consider the evidence in the way the question about truth does. Anger formation involves a further process, which makes anger no longer transparent in an analogous way. Owens' example can be taken simply as bringing out the disanalogy between anger and belief, leaving the argument about belief just as compelling. [39]

I think there is a more promising way to answer the Argument from Transparency, which focuses directly on belief. To reach the word "impossible" in the conclusion, it is crucial that the premises be understood as necessary truths, rather than descriptions of what usually does or ought to happen—norms in either the statistical or the prescriptive sense. As I will now argue, treating the premise as a necessary truth begs the question against the possibility of akratic belief.

A question about belief is not transparent to a corresponding question about the world, when the question about belief is not in the first-person present tense. Does my friend believe it is raining? Did I believe it yesterday? To answer these questions, it is not enough to recognize that the weather reports and the sound of falling drops on the roof present a compelling case for rain. We do not attribute beliefs in these cases based only on what we think is true. To attribute the belief in rain, at the very least, we still need to know whether the person herself recognizes the evidence. We sometimes attribute beliefs that we think are mistaken. We attribute them nevertheless, based largely on observable behavior. If our friend is grieving her husband, she may not notice the sounds of the rain or the weather report on TV. We may see in her demeanour that she is in no position to notice these things. If she mentions that at least it's sunny, we may not know exactly what she is thinking, but, if she seems sincere, we normally attribute the belief to her, even though we

[39] The analogous point about intention is less obvious. One might think that, when asked whether I intend to go swimming, I can answer simply by answering the question whether I am going swimming. This can create the impression that, in the first-person present-tense, questions of intention and questions of truth are settled together. But as even those who see intention as a kind of prediction agree, intending to do something requires more than simply expecting that one will do it. One can expect to go swimming, without intending to go, if one expects to go against one's will, or to change one's mind.

think she should not believe it. Without a general skepticism about the existence of other minds, the default is to continue to make such attributions, and to see them as normally justified.

This kind of "third-personal" belief attribution, in the sense of belief attribution based on observation of behavior, does not require a third person. The attribution of beliefs based on behavior is something we do to ourselves. We sometimes make *third-personal self-attributions* of belief. These are, for example, a key component of various kinds of therapy. Moran (2001, 85) writes:

> Empirically, I can well imagine the accumulated evidence suggesting both that I believe that it's raining, and that it is not in fact raining…. In various familiar therapeutic contexts, for instance, the manner in which the analysand becomes aware of various of her beliefs and other attitudes does not necessarily conform to the Transparency Condition. The person who feels anger at the dead parent for having abandoned her, or who feels betrayed or deprived of something by another child, may only know of this attitude through the eliciting and interpreting of evidence of various kinds. She might become thoroughly convinced, both from the constructions of the analyst, as well as from her own appreciation of the evidence, that this attitude must indeed be attributed to her. And yet, at the same time, when she reflects on the world-directed question itself, whether she has indeed been betrayed by this person, she may find that the answer is no or can't be settled one way or the other. So, transparency fails because she cannot learn of this attitude of hers by reflection on the object of that attitude.[40]

Therapeutic contexts are one kind of counterexample to the necessity of transparency. In therapy, patients often answer the question of what they believe independently of answering the question of what they ought to believe (and independently of answering the question of what is true). This much is true *even if they are mistaken* in attributing these beliefs to themselves, and even if there is no such thing as the kind of unconscious belief with which therapy is sometimes concerned. The unacceptability of particular therapies is of little relevance here. That the patients *attribute* the beliefs to themselves, rightly or wrongly, already shows that we can sometimes settle the question of what we believe in a non-transparent way. This is a key feature of third-personal self-attribution of belief.

According to Heil (1984, 69), this kind of failure of transparency is paradigmatic of akratic—or, as he calls it, incontinent—belief. For him, "The incontinent believer is typified by the psychoanalytic patient who has acquired what might be termed an intellectual grasp of his plight, but whose outlook evidently remains unaffected….He continues to harbor beliefs, desires, and fears that he recognizes to be at odds with his better epistemic judgment." I will not insist that what psychotherapy calls "belief" always

[40] I will not consider Moran's own 'Transparency Condition' here; I make no claim to fit Moran's views or language exactly, but since I quote this passage mainly for the example, I will not go into the details.

counts as belief. But even if it does count, I think it is not the most paradigmatic case of akratic belief, and certainly not the only kind. Therapeutic cases are often on one extreme, in which the attributed beliefs are not (yet) conscious. But there is so far no reason to concede that *akrasia* depends on a belief's being subconscious, preconscious, unconscious, or in any way less than fully conscious.[41]

For a more typical case of *akrasia* in which the beliefs are conscious, consider an example from Adler (2002a, 20).[42] Someone

> suffering from anorexia nervosa can be imagined to be entertaining some thought to the effect that I desperately need to lose weight, but it is evident, as I look in the mirror, that I am thin and do not need to lose weight.

An anorexic typically believes that he desperately needs to lose weight. But if he looks in a mirror, and compares himself to pictures of other familiar and famous people, he might notice that he is thinner than almost everyone he has ever seen. He might then come to believe (here I avoid potentially distracting talk of what is evident) that he does not desperately need to lose weight. But he might also understand (especially after many iterations of this) that he continues to believe that he desperately needs to lose weight. He might recognize his own belief, and might struggle with and resign himself to a belief that the desperate belief will not go away. He would then also see that his answer to the question about his weight does not settle the question of what he believes.

The anorexic can make self-attributions of belief that are third-personal in a further sense than those of some patients in psychotherapy. His self-attributions can be based on observations of his own behavior (like the patient in therapy), and also on observation of

---

[41] Here my line of thought follows Moran (2001), 67:

> From the stance of an empirical spectator one may answer the question of what one believes in a way that makes no essential reference to the truth of the belief, but is treated as more or less a purely psychological question about a certain person, as one may inquire into the beliefs of someone else. If I have reason to believe that some attitude of mine is *not* 'up to me' in this sense, that is, for example, some anger or fear persisting independently of my sense of any reasons supporting it, then I cannot take the question regarding my attitude to be transparent to a corresponding question regarding what it is directed upon. Transparency in such situations is more of an achievement that something with a logical guarantee.
>
> The clash between these two perspectives on oneself is most clearly exemplified in such phenomena as *akrasia*, self-deception, and other conditions where there is a split between an attitude I have reason to attribute to myself, and what attitude my reflection on my situation brings me to endorse or identify with. In such a situation, someone may have good theoretical reason to ascribe an attitude to himself that he cannot become aware of in a way that reflects the Transparency Condition. It may require his best resources of theory and experience to learn what he thinks or feels about something.

[42] I have changed the gender in Adler's example from female to male, to avoid perpetuating the stereotype of anorexics as female.

his own recurring thoughts and feelings (unlike some patients). He can notice these by introspection, and not always by observing his own behavior. The recurring thoughts and feelings he notices by introspection may be signs that he interprets as he would if he observed them in his own behavior, just as he would if they were reported by someone else. He might gather evidence of his own beliefs and other psychological states, both through behavior and through introspection. When his self-attribution is based on this kind of evidence of his own psychological states, rather than evidence about his weight, it is still, in an extended sense, third-personal. And whether he uses behavioral evidence, introspection, or both, the beliefs he has evidence of may be beliefs he believes he should not have.

If it is impossible for an anorexic to be akratic in this way, transparency does not explain why. Adler goes on: "The [anorexic's] thought seems an instance of akratic believing, yet, he does have the thought and so, trivially, it is possible. But what is not possible…is that he cannot *attend* to both conjuncts simultaneously"(2002a, 20). There are two ways to understand what the anorexic cannot attend to simultaneously. One pair of conjuncts is: "I desperately need to lose weight; I should not believe I desperately need to lose weight." Another is: "I believe I desperately need to lose weight; I should not believe I desperately need to lose weight." These correspond to the expressive and reportive-expressive versions of belief-akratic-paradoxical belief in the previous section. The first pair is closer to Adler's example, in which the anorexic's thought includes "I desperately need to lose weight." But only the second pair contains the anorexic's answer to the question of what he believes. The first pair concerns what is true independently of belief, and what he should believe. The second pair concerns what he believes and what he should believe. We so far have no reason to think that he cannot attend to both of *those* conjuncts simultaneously. He might, like the patient in therapy, see that he has a belief that he himself thinks is unjustified. Even fully self-aware or "clear-eyed" *akrasia* in belief is not ruled out here. When someone says: "Of course the evidence shows that flying is not particularly dangerous—certainly less dangerous than driving comparable distances, but I just can't shake the belief that if I fly, my plane will crash and I will die," he reports a belief that, according to him, is unjustified. To think that, in his case, the question of what he believes must be settled in the same way as the question of what he ought to believe, is both to fly in the face of what he says, and to beg the question against the possibility of akratic belief.

Wittgenstein wrote: "One can mistrust one's senses, but not one's own belief"(1953, 190). According to Moran, "This must mean…that neither trust nor mistrust has any application here"(2001, 75). If it is impossible to mistrust one's own belief, we may not be able to speak of trusting it, either. But unless transparency holds without exception, there is one sense in which we can sometimes mistrust our own belief.[43] As

[43] If Wittgenstein's remark is taken as saying that belief is not in general the kind of thing we can trust or mistrust, he would be pointing out a category mistake, rather than insisting on something approaching the impossibility of akratic belief. His thought may be that what we trust or mistrust is not beliefs, but something else—the senses, or reports, or perhaps people. I mean to comment here

Moran points out, "It is a fully empirical question for me whether my own senses or another person's beliefs reveal the facts as they are"(2001, 76). When my relation to my beliefs is partly like my relation to someone else's, I may be able to ask an empirical question about whether they reveal the facts as they are. If I come to believe through observation of myself, either of my outward behavior or of my own recurring thoughts and impulses, that I believe my father betrayed me, my next step may be to collect more evidence. I might come back from therapy, from confession, or from an introspective walk and call my mother to ask her what he really did. If I do this often, I might come to think my third-personally discovered beliefs are more or less reliable—more so, perhaps, about my father, but terribly off about my sister. A "normative ideal" of transparency, as Moran puts it (2001, 62), leaves open the possibility of wild and surprising deviation.[44]

Examples of transparency failure bring out that transparency is most properly said to be present or lacking not in belief formation or retention, but in belief *attribution*. It is in answering the question whether we have some belief that we normally look to evidence of its truth. The transparency of a belief, or of a question about whether one has some belief, then has only indirect bearing on whether we can actually form or retain a belief that we

---

only on a particular interpretation of Wittgenstein's remark, without insisting that it is what he had in mind.

[44] As may by now be apparent, such examples also offer an explanation of the possibility of Moorean belief. Through therapy, I may come to believe: "I believe my father betrayed me, but he didn't." Third-personal self-attribution thus also provides an argument against premise (1) of the Argument from Moore's Paradox, which I conceded in the previous section.

Perhaps more surprisingly, consideration of third-personal self-attribution suggests a way in which some Moorean and belief-akratic-paradoxical beliefs may be not only possible, but rational. Convinced by his doctor, an anorexic may conclude: "I do not need to lose weight." But he may also acknowledge the persistence of his belief that he does need to lose weight, based on observation of his own continued and often extreme emotion and behavior. His belief that he does need to lose weight may well be irrational; but the belief that he has that belief, and that it is false or that he should not have it, can itself be a rational one. The 'anorexic' belief may the product of insecurity and a warped body image; but the Moorean and belief-akratic-paradoxical beliefs themselves can indicate an impressive and hard-won self-awareness.

Indeed, for Moorean and belief-akratic-paradoxical beliefs to be rational, the third-personal self-attribution does not even have to be correct. A psychotherapist, a friend, or an observer may, through incompetence or malice, leave me misinformed about the details and implications of my current mental life. I may be told that I yell "Traitor!" at my father every night while asleep. Or I may be misled into accepting that my anger and loneliness are reliable signs of belief in betrayal. With enough negligence or intrigue on the part of others, I may justifiably come to believe that I believe my father betrayed me, even as I continue to believe that he in fact did not, and that any such belief about him is grossly unfair. I can then have the Moorean belief: "I believe my father betrayed me, but he didn't," and the belief-akratic-paradoxical belief: "I believe my father betrayed me, but I shouldn't believe it." Although my having these beliefs is still troubling, in such cases the fault can lie with someone other than myself. I may thus be able to falsely but justifiably attribute to myself a belief that I believe is false and *un*justified, again resulting in rational Moorean and belief-akratic-paradoxical beliefs.

believe we should not have.[45] Whatever this bearing is, the appeal to transparency in this context assumes a common feature of belief to be a necessary one. That assumption is false. Transparency sometimes fails. The akratic anorexic would be an example of its failure. The assumption of the necessity of transparency is an unlikely one, and rules out some forms of akratic belief in advance. An emphasis on transparency is thus another way of expressing a deep puzzlement about the possibility of akratic belief. It is not an independently compelling way of ruling out that possibility.

In considering these impossibility arguments, I have not given an exhaustive list of possible denials. Nor have I given an exhaustive classification of views about the nature of belief, and shown that on each of them, there is no reason to deny the possibility of beliefs we believe we should not have. I have tried to develop and then answer the thoughts that are most natural and most often expressed, even if in passing, in denials of the possibility of akratic belief. These answers also suggest a more compelling way of addressing the denials. The denials can arise in various forms, of which, if I am right, the most natural ones can be shown to have no independent argumentative force. But they do express an underlying puzzlement that has not yet been addressed. To address that puzzlement, we need a recognizable characterization of akratic belief that undercuts the motivation for denying its possibility.

[45] One might think that, even if the question whether I believe it is raining is not transparent to the question whether I should believe it, the question whether *to* believe it is raining *is* transparent in this way. That question is more directly relevant to the formation and retention of belief. But here the same conclusion can be reached in a slightly different way. Consider the belief that my father betrayed me. Even if my answer to the question whether to believe this depends entirely on my answer to the question whether I should believe it, a negative answer might not stop me from having the belief. There is still the possibility that I can decide not to believe something, and nevertheless continue to believe it.

Shah and Velleman (2005) offer a systematic treatment of the transparency of the question whether *to* believe. What I say is in line with their view, which insists that belief can be formed independently of, and be unresponsive to, our views of what we should believe. "One may reason one's way to the conclusion that one's plane is not going to crash, for example, and yet find oneself still believing that it will…. In this case, an irrational phobia has had a dominant hand in determining what one believes."(507-8). They even accept that in such cases, "one is in a position to have a thought with the form of Moore's paradox: 'The plane will be safe, but I don't believe it'"(508).

Chapter 3: Believing Against One's Better Judgment, II: How Akratic Belief is Possible

So far I have tried to show that there is no principled reason to think that akratic belief is impossible. The arguments against the possibility of akratic belief have all expressed a sense that there is no room for believing something that we believe we should not believe. They have all been ways of drawing out that idea. Each of the arguments I answered tried to explain why the presence of a 'better judgment' rules out the possibility of someone's believing akratically. According to them, our picture of that person's mind leaves no room for the conflicting belief.

Though I addressed some particular arguments, I have not yet addressed the underlying doubt. How can we believe something if we, right now, believe we should not believe it? I think the doubt remains, because it is still hard to wrap our minds around the phenomenon. The sense of paradox is not completely removed by answering the impossibility arguments. With or without a refutation, it is hard to see how akratic belief is possible.

In this chapter, I address the underlying doubt by developing a conception of akratic belief. To do this, I will combine elements of two earlier accounts by Amélie Rorty and T.M. Scanlon. Drawing on their views of akratic belief, I will consider several key marks of belief, based on which belief is commonly attributed: sensitivity to evidence, recall in relevant circumstances, conviction, reporting, and use in further reasoning. I will argue that in akratic belief, both component beliefs in the akratic state manifest these characteristics to an extent we normally recognize as belief, while nevertheless conflicting with and partly undermining each other.

A natural way to defend the possibility of akratic belief is to defend a theoretical conception of belief, and then show how that conception allows belief to be akratic. But any such line of defense ties the explanation of akratic belief to the success of a particular theory. There is a more ambitious line of thought to be pursued here. If I am right, the possibility of akratic belief should be treated like the possibility of akratic action often is. It should be accepted as a pre-theoretical datum which a conception of belief should be able to accommodate. That is, akratic belief should be seen as a puzzling but recognizable phenomenon with wide-ranging theoretical implications. This is the position I will defend for it. I will thus avoid, as much as possible, relying on any particular theoretical view about the nature of belief, though I will come back to the question of the extent to which one can be neutral with respect to those theories. I will try to show how, on a wide variety of views about belief, we can both recognize a belief as akratic, and understand why akratic belief is puzzling.

I. Rorty's Catalogue

According to Rorty (1983), *akrasia* in belief is not only possible, but common and widely varied. Her article "Akratic Believers" presents "a catalogue…locating possible akratic breaks"(177). She describes four distinct types of *akrasia* in belief, which she calls "intellectual," "interpretative," "inferential," and "practical."

In Type 1, or "intellectual," *akrasia*, a person "fail[s] to commit himself to his general beliefs about what is best, divinely commanded or morally desirable," either "refusing" to follow them or "voluntarily failing." The first of these includes Milton's Satan proclaiming: "Evil be thou my good." The second includes allowing oneself to succumb to depression.

Type 2, or "interpretative," *akrasia*, occurs "between a person's principles and commitments on the one hand and his interpretations of the situation in which he finds himself on the other"(177). Rorty subdivides this type into *akrasia* of perception, of description, and of emotion. Perception, broadly speaking, can fail to conform to our views of what ought to be salient. "Someone who denies ageism might see the lines on the face of the elderly as deformations, their motions as comical… A painter who has become a military commander might akratically look at a landscape as a composition"(177). The phrases we use to describe a situation can similarly fall into ways of talking that we disapprove of, leading to *akrasia* of description. In Rorty's main example, "someone committed to non-sexist attitudes…talks of women as *broads* or *chicks*…. What he calls imaginative initiative in a man, he calls conniving manipulation in a woman"(178). Emotional reactions can follow a similar pattern. "A person might for instance be hostile to someone whom he believes to be friendly, knowing that he does so solely because of a superficial resemblance to an ancient enemy"(178). In all of these varieties, the way we intepret a situation is a way we are committed to not interpreting it.

In Type 3, or "inferential," *akrasia*, a person "come[s] to a conclusion following a pattern of inference that he regards as illicit"(179). We can, for example, accept a view for the sake of argument, knowing that this will lead us to accept it, period. Or we can conduct an inquiry in ways that will "predictably confirm" our hypotheses. "A whole scientific community, or a governmental elite… can follow habitual and comfortable procedures that they do not underwrite or that they regard as irresponsible modes of investigation"(179).

Type 4, or "practical," *akrasia*, Rorty also calls "akrasia of intention and decision. It occurs when the conclusion of a piece of practical reason fails to conform to the premises. This sort of akrasia stands halfway between akrasia of inference and akrasia of action"(179). Being akratic in this way is making practical decisions in ways one does not endorse: for example, "allowing daydreams to have more weight than [one] thinks they should," or forming a comparative resolution contrary to the balanced outcome of practical reasoning. "Few" of us, Rorty says, "can…change what they consider inappropriate patterns" of practical thought (180).

Rorty's "catalogue" of "possible akratic breaks" can be summarized this way:

1: Principles || commitments
2: Principles, commitments || interpretations of the situation
3. Principles, commitments, interpretations of the situation || inferences
4. Principles, commitments, interpretations of the situation, inferences || decisions

More succinctly, the breaks can be lined up as follows:

Breaks: principles |1| commitments |2| interpretations |3| inferences |4| decisions

Rorty's catalogue can be seen here to be not just a haphazard assortment. Rorty has traced the process of inquiry from general principles to the decisions in which inquiry concludes. She gives vivid examples at every stage, helping to distinguish them from each other and attempting to show in detail how belief can be akratic.

The catalogue has a problematic feature, which runs through all four types. All the types and most of the examples seem to concern not belief, but the voluntary activities that result in a belief. Failure to commit to one's principles (Type 1) is either a "refusal" or a "voluntary failure." 'Seeing' wrinkles as deformations or a landscape as a composition (Type 2) is, if not an action, at least the taking up of an attitude other than belief, a different way of regarding or thinking about something. Conducting an inquiry (Type 3) is itself an activity. Making decisions in ways one does not endorse (Type 4) sounds like a practical failing; or, at least, we would want to know more about why the concusion does not conform to the premises before it is clear how the *akrasia* is only "halfway" to *akrasia* of action. There is a surprisingly strong element of the voluntary throughout Rorty's catalogue of akratic 'breaks'.

That element is explicit. Taking action as her paradigm case, Rorty assumes that *akrasia* of any kind must involve an element of the voluntary. Her treatment of belief turns on "the central issue of whether beliefs and varieties [of] intellectual actions that form them can be voluntary"(176). She "argues…for treating believing as the sort of voluntary condition that can be akratic"(181). And her considered view is that "Since akrasia of belief has the same structure as akrasia of action, some kinds of believings are, for some kinds of people, as voluntary as some kinds of actions are, for some people"(175). Rorty takes voluntary failure to be essential to *akrasia*. So in giving an account of akratic belief, she tries to show how beliefs (and the processes that form them) can be voluntary. Her catalogue is "a catalogue…of voluntary beliefs, locating possible akratic breaks"(177).

It is natural to feel that Rorty is cheating. We are trying to understand how someone can believe what they believe they should not believe. But voluntary failures seem to be paradigmatic cases of *akrasia* in the *practical* realm—standard cases of akratic action and intention, only secondarily related to belief. Voluntarily letting oneself lapse in commitment to a principle, or engaging in inappropriate intellectual inquiry, is not clearly an example of having a belief one believes one should not have. In some of Rorty's cases it

might be hard to make out any belief at all. And it can seem obscure how any of them shed light on the phenomenon of believing against one's own better judgment.[46]

There is a motivation for thinking that akratic belief must be voluntary. In the case of action, it is usually thought that *akrasia* must be distinct from mere compulsion. Aristotle (1999, III.1-5 and VII) thought that *akrasia* must be voluntary and so in a sense "up to us"; and we normally think of action against our better judgment as something that, unlike movement caused by an irresistible urge, we are responsible for. To akratically eat a dessert is to eat it against one's better judgment, but to nevertheless have control over one's own movements. That is a central part of the puzzle of *akrasia*; akratic movement seems paradoxically ours and at the same time not ours. Rorty can be seen as trying to

---

[46] Since Rorty's catalogue has several subdivisions and a variety of examples, the problem is intricate, and the doubt that they capture *akrasia* in belief will take different forms. I do not go through all of them in the text. Distinguishing the ways in which parts of the catalogue may be "cheating," or describing what only illicitly appears to be akratic belief, gives rise to a countercatalogue of pitfalls to avoid in describing the variety of akratic belief:

(1) *No belief*: Describing an example of akratic belief in which the akratic state does not include a belief. Though this might not seem worth mentioning, it is prominent in Rorty. Under her Type 1, refusing or voluntarily failing to follow a principle does not seem to require believing that the principle is not binding. Under Type 2, "interpretative *akrasia*," "emotional" akratic states such as hostility may be loosely described as interpretations of a situation, and the latter can be loosely associated with belief. But hostility is not itself a belief. Neither is a sexist choice of words, whether deliberate or a slip. Other cases are less clear cut: seeing the motions of the elderly as comical may or may not be believing that the motions are comical.

(2) *Mere result*: Calling a state akratic because it is caused by an akratic action. An extreme case would be taking a pill that paralyzes you, or gives you a new, coherent set of beliefs which you would now reject, or makes you think Pluto is still considered a planet. Even if you take the pill against your better judgment, we would not say you are then akratically paralyzed, or that hold your new beliefs akratically. Rorty's Type 3, "inferential" *akrasia*, fits this pattern. Apart from pulling attention away from belief, the description entails only pseudo-*akrasia*: it appeals to a 'better judgment' that may be held at a different time than the state in question. Worse still, the judgment is not about that state.

(3) *Mental action*: Assuming that an akratic state is not *akrasia* of action, because it involves no observable bodily movement. Though this is not explicit in Rorty, it can be part of the underlying appeal of her catalogue. When seeing a landscape as a composition, accepting a view merely for the sake of argument, or reasoning irresponsibly, it can seem that the *akrasia* must be of a special kind because there is no action there to speak of. But lack of bodily movement shows little; doing long division in one's head is not recognizably less of an action than doing it on paper.

(4) *Underdescription*: calling a state *akrasia* of belief when it is not yet clear what kind of failing is at issue. Mere Result (above) can be seen as a kind of underdescription, since the result itself is underdescribed. The pitfall may be clearest in Type 4, "akrasia of intention and decision," which "occurs when the conclusion of a piece of practical reason fails to conform to the premises." A conclusion can fail to conform to premises in many different ways, among them a simple mistake. Rorty rightly says only that these cases are "sometimes" akratic (179). But the underdescription makes the description fall short of singling out *akrasia*. It then picks out not a type of *akrasia*, but an area of life (compare friendship, sex, or sports) in which *akrasia* can occur. We then make little progress on the larger underdescription problem of making clear how a belief can count as akratic.

capture the parallel puzzling phenomenon in the case of belief. Her view also fits well with the etymology of "*akrasia*" as lack of self-control. What she leaves out is the possibility that akratic belief can require a different kind of control, or that no notion of control is needed to understand it.

I will soon return to this issue. What I want to point out in Rorty is what is left even if we concede that talk of the voluntary is unnecessary and even distracting in thinking about akratic belief. Rorty sees herself as teasing apart different elements that normally go together in belief. As she concludes from her catalogue, "The phenomena standardly classified together as *believing* are in fact quite various and diverse," including "attending, focusing, seeing as, classifying, describing as"(181). Rorty's catalogue exemplifies a method which might be summed up by her phrase: "distinguish the strands"(181). Rorty tries to distinguish the strands of belief and of reasoning to and from a belief. When we see how these different strands or elements or components can come apart, we may be able to see how they might go against each other, exhibiting both a belief and a second belief that one should not have the first one. If most of Rorty's examples are too standardly examples of *akrasia* in action or intention, we can hope for a different set of strands that more clearly pertains to belief. To develop that set, I will draw inspiration from T.M. Scanlon.

II. Scanlon's Dispositional View

Scanlon's (1998) consideration of akratic belief begins by raising a doubt about its possibility. There can seem to be, as he too puts it, "no room" between our beliefs and our view of the reasons for them. It can seem that "To take P to be supported by the best evidence just *is* to believe it"(35).[47] "But this," he says (35),

[47] The opposing view as Scanlon expresses it goes nicely with the denials of the possibility of akratic belief considered in the previous chapter, all of which insist that there is "no room" to form an akratic belief once we form a view of the evidence or reasons against it. But it is interesting that Scanlon's version does not match any of the ones I considered. It is not a nullification argument, or a Moorean paradox, or a transparency argument. It is not exactly an argument at all, but, it seems, a view that identifies a belief in conclusive evidence for a proposition with belief in the proposition. It is not a very common or attractive view. Few of us would say that our belief that grass is green is *the same as* our taking the best evidence to support that grass is green. As far as I know, this view has never been appealed to or expressed in arguments against the possibility of akratic belief. For that reason, I leave out this unlikely view in Chapter 2. And I am inclined to read Scanlon charitably as reading the denials themselves charitably. That is, he might mean that, on that view, to take P to be supported by the best evidence is *thereby* to believe it, in the sense that anyone who does the first must do the second. In that case, he is giving a dramatic way of re-expressing the "no room for slippage" idea, rather than pointing to a particular, uncommon and extreme view about belief as a central source of support for it.

It is also interesting that Scanlon here ignores the possibility of contradiction in belief. If to take P to be supported by the best evidence just *is* (thereby) to believe it, it is still entirely possible to akratically believe not-P. One just has to believe P and *also* believe not-P. Scanlon's taking the view he describes as a threat to the possibility of *akrasia* is one example of the widespread neglect of the possibility of conflicting states in the literature on the topic. For this reason too, if that view

> seems to me a mistake. Belief is not just a matter of judgment but of the connections, over time, between this judgment and dispositions to feel conviction, to recall as relevant, to employ as a premise in further reasoning, and so on. Insofar as akrasia involves the failure of these connections, it can occur in the case of belief as well as in that of intention and action. I may know, for example, that despite Jones's pretensions to be a loyal friend, he is merely an artful deceiver. Yet when I am with him I may find the appearance of warmth and friendship so affecting that I find myself thinking, although I know better, that he can be relied on after all.

To believe, according to Scanlon, is in part to be disposed to feel conviction, recall, and employ the belief in a range of situations. *Akrasia* involves "a failure of these connections."[48] We might know (or judge, or believe) our friend Jones to be a skilled deceiver and his friendship to be a con. But the connections fail. In his presence we are no longer so sure he is a deceiver; the evidence against him may not come to mind as relevant; and we do not draw the further conclusion that, for example, we should not lend him our money or our loved ones. We lose sight of the artifice; the results of reflection waver in the face of such skilled deceit.

As we saw in Descartes, one can doubt that such a description captures a case of *akrasia*, rather than a pseudo-akratic forgetting or change of mind. Scanlon's description comes closest to claiming *akrasia* in the phrase: "I find myself thinking, although I know better." A doubt can be raised about both the thought and the knowledge. "I find myself thinking" is ambiguous between "I find myself believing" and "I find the thought occurring to me." In the second of these, I may find myself, for example, *inclined* or *tempted* to believe that Jones can be relied on after all, though I am sure he is not. Or I might think: "Hmm. Maybe he can be relied on after all?" And then I am no longer sure. To have an akratic belief, I must actually have a belief. The passive-sounding "find myself" suggests a belief to which I stand in a particularly passive relation. But "find myself thinking" also suggests that the state may be something other than a belief. Whether it can be a genuine belief is precisely what is at issue.

On the other hand, suppose I do find myself *believing* that Jones can be relied on. Do I still "know better"? Here Scanlon's view can seem self-defeating. Scanlon says:

---

were taken as an argument against the possibility of akratic belief, the argument would not be a very good one. I consider conflicting beliefs in more detail in the next chapter.

[48] Scanlon here distinguishes between judgment and belief. In the passage from Scanlon, as in ordinary language, it is not immediately clear what the difference between them is. Is what is at issue a difference between event and state, between conscious or self-conscious thought and thought that may or may not be self-conscious, or something else? What is judgment if it does not itself involve dispositions to feel conviction, recall as relevant, and so on? In this chapter I avoid the issue of the relation between judgment and belief, since my aim is to account for *akrasia* in belief. But a view of judgment as a state or 'act' distinct from belief may be able to draw on my account to explain *akrasia* in judgment. For more on the distinction between judgment and belief, see Cassam (2010).

"Belief is not just a matter of judgment but of the connections, over time, between this judgment and dispositions to feel conviction, to recall as relevant, to employ as a premise in further reasoning, and so on." If Scanlon is right, this also holds of the belief that Jones is an artful deceiver, and of the belief that I should not believe that Jones can be relied on. But these beliefs, it seems, do *not* preserve their connections over time. Drawn in by Jones's charms, I presumably do *not* feel conviction in, do not recall as relevant, and do not reason from those beliefs. Instead, I feel more and more sure that Jones can be relied on after all, and I reason from that. So it can seem unclear how, on Scanlon's view, I count as 'knowing better', or believing that I should not believe that Jones can be relied on. This again gives the appearance of mere pseudo-*akrasia*, an irrational change of mind rather than a belief that I believe I should not have. I seem not akratic but fooled.

Nevertheless, I think Scanlon has the seeds of the right answer. To see this, we can look more closely at his guiding idea, and try to expand it into a catalogue like Rorty's.

Rorty's aim was to "distinguish the strands" in belief and intention, showing that "The phenomena standardly classified together as *believing* are in fact quite various and diverse," and cataloguing the ways in which *akrasia* can arise in order to understand how belief can be akratic. Her list of the phenomena standardly classified together was "attending, focusing, seeing as, classifying, describing as," and, as we saw from her catalogue, committing, perceiving, interpreting, inferring, and concluding.

Scanlon offers a similar distinguishing of strands, but with a different list.[49] Scanlon's list is: "dispositions to feel conviction, to recall as relevant, to employ as a premise in further reasoning, and so on." We can say that to have a belief is, in part, to have a sense of conviction in it, to recall it at appropriate times, and to reason from it. But Scanlon's view includes three further elements. First, what is central are *dispositions*, not an actual recalling or feeling of conviction. Second, he thinks belief is a matter of the *connections over time* between the judgment and the dispositions. Third, he says "and so on," leaving room for additions to the list. For him, belief is partly a matter of *connections over time* between judgment and *dispositions* to feel conviction, to recall as relevant, to employ as a premise in further reasoning, and possibly something else.

One might object at the start that Scanlon's is the wrong kind of view to make sense of *akrasia*. On Scanlon's view, one might think, belief requires these connections over time. So if the connections are broken—if a person does not recall a belief as relevant, or use it in further reasoning, then the person does not count as having the belief at all. Broken connections show a forgetting or change of mind, so Scanlon's view is at best an account of pseudo-akratic states. To think that Jones can be relied on after all is to fail to

[49] There are still differences here between Rorty's treatment and Scanlon's. For example, Rorty aims to distinguish phenomena that she says are "standardly classified together," while Scanlon's insistence that belief is "not just a matter of judgment" suggests that he thinks the other phenomena are ignored, or kept too separate, rather than conflated. The key parallel I want to point to is the distinguishing of several closely related elements. Here Scanlon's thought is in one way closer to mine than Rorty's is. His talk of what belief is 'a matter of' suggests that he is more concerned than Rorty to insist that the various elements are ones by which belief itself is properly attributed.

use one's belief against him in further reasoning (and probably to fail to recall it). Since believing that Jones is an artful deceiver is a matter of connections with events like these, one does not really know better, and so is not akratic. Scanlon's view seems to require too much of belief, ruling out the presence of at least one of the beliefs in the akratic situations we want to account for.

This objection is not hard to answer. According to Scanlon, belief is a matter of connections over time with a range of *dispositions*. Nothing has shown that someone lacks any disposition with respect to the belief that Jones is an artful deceiver. When Jones is not around, we may be brooding on his deceit, feeling sure of it, working out its implications for his character, and even resenting him and planning to avoid him. Jones' presence can work to effectively block the manifestation of dispositions we do have. His charm, his concern about our welfare, and his innocent-looking face can make it harder to recall that belief or reason from it. It can do this without removing the underlying dispositions. This is part of the way we understand the difference between a disposition and its manifestation. Fragile glasses do not always break, even when hit or dropped.

The objection does bring out that Scanlon's view cannot require *all* of the dispositions to be readily manifested at all times. (That, one might think, really would rule out *akrasia*.) But because his view involves a variety of dispositions, it can allow some degree of systematic failure of or interference with those dispositions. We can then ask what kind of and how much failure constitute *akrasia*, and what kind or how much constitute a lack of belief.

Another, more serious problem can be raised at the start. Scanlon expresses a quite particular, dispositional view of the nature of belief. Whatever its implications, can it be the right explanation of how belief can be akratic? It would be unfortunate if the explanation depended on a controversial view of belief, itself not defended on other grounds. The explanation would not be acceptable to most of us, who hold either a different general view of belief or none at all. And one might doubt that such an explanation could be right in principle. One might think that a resolution to the sense of puzzlement about akratic belief should be addressed to everyone who is puzzled.

To avoid this problem, I will avoid relying on Scanlon's view. Dispositional views have able defenders[50], but I will try to stay as neutral as possible between competing views of belief, since the goal is to explain in general terms how *akrasia* is possible, without tying the explanation to a narrow range of views in the philosophy of mind or epistemology. To that extent even Scanlon's dispositional view is meant here as an example of the more general strategy of "distinguishing the strands" in belief (as he may himself intend it to be in using the vague phrase "a matter of"). I will try to develop that strategy by describing several widely recognized marks by which belief is attributed, without defending a view about the metaphysics of belief and its relation to these 'strands'.

To develop the strategy, we can add two more elements to Scanlon's open-ended list. One further element that is often thought central to the attribution of belief is the

[50] See Schwitzgebel (2002; 2010), discussed below, for a recent defense and discussion with references to earlier dispositional views.

reporting of the belief. If asked: "Do you believe Jones is a deceiver?", it seems that someone who does have the belief would at least sometimes say yes. Second, we usually think beliefs must be sensitive to (apparent) reasons for them. If we believe that smoking is harmless, we will have at least some disposition to reconsider that belief when we read studies of its effects. If we believe that Jones is an artful deceiver, we will be ready to consider, in this case with some suspicion, evidence of his extraordinary honesty, or to redouble our confidence when we hear the stories of his other victims. Someone who does not think these elements are central to the attribution of belief can take the additions themselves with a grain of salt, and adapt what I go on to say accordingly. But for now I take the liberty of making these additions to Scanlon's list where he writes "and so on." Belief is often attributed on the basis of sensitivity to reasons, recall in relevant circumstances, conviction, reporting, and use in further reasoning.

What happens when these marks are absent? We can distinguish a different kind of failure corresponding to each of them.

(1) *Dogmatism*. Some beliefs show little or no sensitivity to evidence. This can be true of some astrological beliefs, like the belief that people born in August are more courageous than people born in March. We can imagine it true of the belief that Jones is a loyal friend, before we finally faced the evidence of his deceit.
(2) *Lack of recall*. We can have beliefs but fail to recall them. We might learn that the door of a room opens in rather than out, but need many visits before we remember this fact in time and stop pushing before pulling. A belief in Jones' deceit may not come up as relevant while he is peppering his moving story of how badly he needs money with casual mentions of how good he is at paying it back.
(3) *Lack of conviction*. A depressed person can agree that life is worth living; Rorty's professed anti-sexist can agree that women and men are equal in intelligence and instrinsic worth. But they might not feel much conviction in these beliefs. They can experience themselves as just "going through the motions." If we recall Jones' deceit in his presence, we might even refrain from lending him money, but without a feeling of conviction in the belief we act on.
(4) *Denial*. When asked if they believe their career might fail, or have ever been a victim of abuse, or would support Stalin's regime, many people say no. Some of them are truthful; others lie; others might not be able to face the fact that, when it comes down to it, they do believe that their career might fail, or that they have been victims of abuse, or that, in fact, they do support Stalin's regime. Similarly, we might not report believing that Jones is an artful deceiver (or a loyal friend), for various reasons.
(5) *No further reasoning*. We can acquire a belief and still have trouble reasoning from it. Physics students often make mistaken predictions about angular momentum, sticking to misleading intuitions even after learning its laws. I can believe that my friend Jones is an artful deceiver, but not draw the conclusion that, for example, it might be worth asking myself why he tells me his moving stories the way he does.

Or when he asks for my trust, I might think "He's a deceiver!", but still go along with him.

With all of these failings combined, we lose our sense that a person has the belief in question. Suppose I believe that a Democrat should be President of the United States. But I do not recall the belief, I feel no conviction in it, I do not report the belief when asked, I do not reconsider it when faced with evidence for or against it, and I vote Republican. At this point, it seems unclear how I count as having the belief at all.

But we can recognize a belief when one or more of these failings are present. Belief survives some systematic failings along these lines. All five particular failings have examples that seem to still count as belief, including, I think, the examples I gave above. And belief can survive more than one failing. In a conservative town, I might be reactionary and secretive, thinking and researching and voting Democrat, but refusing to tell others or consider the merits of their position. Or I might declare with conviction that I am a Democrat, although I keep finding "special" reasons to vote Republican in particular cases, tend not to recall that I see myself as a Democrat, and am strikingly insensitive to this and other evidence that I am not as Democratic as I think. The type and degree of failing that belief allows is a matter of controversy. But it is not hard to agree that there is leeway here.

On the other hand, it should be clear from this second catalogue that not all of the failings are akratic. A secret or dogmatic belief is not thereby an akratic one. To the extent that the list is a catalogue of failings, it is not a catalogue of varieties of *akrasia*. How, then, does it help?

III. How Can Belief be Akratic?

When do these various failings amount to *akrasia*? An akratic belief is a belief one believes one should not have. Scanlon's view of belief—and any other view of belief—should apply to *both* of those beliefs. If Scanlon is right, the failings are akratic when they involve a second, competing judgment, together with connections over time to the corresponding dispositions, to a high enough degree to count as a belief. Though both beliefs will have failings on some of the dimensions in the Scanlonian catalogue, we can also use the failings to see how the beliefs are related to each other.

Take "Jones is a loyal friend" and "I should not believe that Jones is a loyal friend." If we have both of these beliefs, the first will not be ideally sensitive to evidence. If it were, we would reconsider it in light of the evidence I see of Jones' deceit. On the other hand, the second belief will not be one we apply well in further reasoning. If we did reason from it, we would probably come to at least suspend the first belief. In practice, we probably will not even recall the second belief in some circumstances. When Jones is around, the second belief may be far from our minds, even though we angrily fixate on his lack of loyalty at other times. Our feeling of conviction in both beliefs might waver, as will

our reports, though it might be natural to expect that we report the second belief more often. We then have a relatively dogmatic belief that Jones is a loyal friend, and a less than ideally operative belief that we should not believe this.

The same can be said of the examples in the previous chapter. In *Fear of Flying*, Matt can believe: "My plane will crash," and also: "I should not believe that my plane will crash." The first belief is hardly sensitive to evidence, though Matt does gesture at a justification by asking: "What's holding it up there anyway?" He feels some conviction in the first belief—at least enough to sustain the terror—and also in the second, since he confidently proclaims his belief about evidence, although he fails to reason from it when the opportunity to fly arises. He seems to recall and report both beliefs.

What would it take for Descartes and Hume, or at least their readable incarnations, to be akratic? They might believe: "There is a tree in the garden", and "I should not believe that there is a tree in the garden." They would then—at one stage in their respective books—have the second belief that they should "reject" or "withhold assent" from the first. But they would find their conviction returning to the first belief, and their reasoning guided by it as they walk around the tree. Their case is tricky, since it involves skeptical reflection and differing levels of abstraction, which complicates the case psychologically for them and descriptively for us. But we can see roughly how it might go. Their reflective view of the evidence supports the second belief, and (we can imagine) they continue to report it, feel some limited conviction in it, and perhaps reason from it to some extent. The first belief, that there is a tree there, will also carry some conviction, may also be reported in conversational contexts, and is likely to be reasoned from in, for example, walking around the tree. We can even think of it as responsive to reasons, since the tree (or the fact or the seeing of it) presumably counts as an apparent reason for the belief. There will be a wide range here as the course of a day shifts the focus, with the thinker holding on to one belief and trying to forget or give up the other. But the reasoning, reports, conviction, recall, and (at least apparent) reasons on both sides make the example in one way easier. It is easier to see how there can be two beliefs here, even as the beliefs oppose each other. Put Matt on a plane, me with Jones, and Hume at a backgammon table, and we might all insist on the reality of the table, friendship, or impending crash. In other contexts, we insist that we should not have these beliefs. The distinctive context in which each belief shows itself helps it show its colors as a genuine belief. Despite their failings, both beliefs maintain core marks of belief, even as they come into conflict with each other.

The thought of "core marks of belief" can suggest that we are working with a particular theory of belief, which might be only one candidate theory among many others. But it is worth emphasizing that, in several ways, the kind of explanation I have given is neutral with respect to theories of the nature of belief, in the sense that it allows a wide range of theories of belief to account for the possibility of believing akratically.

First, as we saw, Scanlon's dispositional view, even with an expanded list of dispositions, is only one example of a theoretical conception of the nature of belief. One can think of belief as a different kind of disposition, or as a representation stored in the mind or brain, or as a relation to such a representation, or as a physical state of the brain, or

as a functional state, or as some combination of these, or in some other way. Though I have talked about marks of belief, I have not said anything general about what beliefs are.[51]

Second, I have not said whether the five characteristic marks of belief should be thought of simply as marks by which one can attribute a belief to oneself or others, or also as partly or entirely *constituting* belief—as themselves being the belief state. For our purposes, it doesn't matter, except insofar as it allows a description of akratic belief to be consistent with a wide range of theoretical views. This second kind of neutrality is useful for the first. Whatever beliefs are, metaphysically speaking, we can focus on the conditions under which they are properly attributed. The relevant marks are the marks used in attribution.

Lastly, I do not insist that the five marks I considered are the most important ones, or that they are all essential or even significant. Even the list of marks is just one example. The description of akratic belief can hold even when the details are different, or limited—even when, on some theoretical views of belief, a "distinguishing of strands" is unnecessary or even impossible. In the Appendix to the *Treatise of Human Nature*, Hume (2000, 396) writes: "belief is nothing but a peculiar feeling, different from the simple conception." Cohen (1992, 5) writes that "Belief is a disposition to feel"; that is, although it is a disposition, "Belief is a disposition normally to feel that things are thus-or-so, not a disposition to say that they are or to act accordingly"(1992, 8). Views like the ones expressed here seem to have only one "strand" in their proper attribution: as Cohen (1992, 11) puts it, "credal feelings"(11), such as feelings of conviction that some proposition is true. If they are right, a defense of the possibility of akratic belief seems especially easy. One need only find *feelings*, or dispositions to feel, both that some *p* is true, and that one ought to believe it. Feelings are notoriously capable of conflict with each other.[52] The attribution of akratic belief thus does not depend essentially on any "distinguishing of strands." Instead, the distinguishing of several marks by which we attribute belief is useful

[51] Influential representationalist views include Dretske (1988), Fodor (1975, 1981), and Millikan (1984). Dispositional views include Braithwaite (1932), Price (1969), Audi (1972), and Schwitzgebel (2002). Influential functionalist views include Armstrong (1968), Lewis (1972, 1980), and Putnam (1975). For a useful recent survey of these and other views, see Schwitzgebel (2006). I do not claim consistency with *all* views about belief; most obviously, those who deny the existence of beliefs altogether, as Churchland (1981) does, will not be likely to accept my explanation of how belief can be akratic. I claim only neutrality across a wide range of theoretical conceptions of belief, and a lack of reliance of any particular theory. If I am right in accepting the possibility of akratic belief as a pre-theoretical datum, that possibility will then count against theoretical conceptions of belief that cannot accommodate it.

[52] Williamson (2000, 99) suggests another one-strand view: "Intuitively, one believes *p* outright when one is willing to use *p* as a premise in practical reasoning." If we again take the remark out of context, it suggests a simple theoretical conception, on which use in further (practical) reasoning is the central mark by which belief is properly attributed. Here again, it is not hard to imagine the two component beliefs in an akratic state used willingly as premises in practical reasoning. If our practical reasoning can itself be inconsistent, it should be able to justify the attribution of akratic belief.

for giving a more detailed picture of akratic conflict, in a way that makes vivid what akratic belief is like. It brings out the ways in which the marks by which we recognize belief can be seen both in an akratic belief and in the 'better judgment' that renders the first belief akratic. The picture of conflicting marks can hold when the marks are fewer or different in kind.

So far, we have seen two lines of thought that support the possibility of akratic belief. First, I gave a series of examples: the anorexic, the superstitious person, the cheated spouse, and the gambler, all stuck, with some degree of self-awareness, believing what they themselves believe they should not believe. These examples seem to be vivid cases of akratic belief, and offer an initial case for its possibility. Second, I offered an argument from belief attribution, arguing that the ordinary marks by which we attribute belief can sometimes be present in both component beliefs of an akratic state. We often look to abnormal cases to help us understand the normal ones, and we can think about akratic belief in this spirit. But within an argument from belief attribution, it is the normal cases that help us understand an abnormal one. When we consider how we attribute beliefs more generally, I argued, we can come to see at least some apparent cases of akratic belief as ones in which the beliefs in question are properly attributed. The appeal to examples and the argument from belief attribution are naturally combined. Some examples of belief seem to be cases of akratic belief; and when we consider the conditions under which we attribute beliefs more generally, we can see that, and how, belief can sometimes be akratic.

At this point, puzzlement about the possibility of *akrasia* can culminate in a final statement of the central doubt. Why think that, in these cases, the 'marks of belief' are sufficiently present? In other words, why think the 'beliefs' count as genuine beliefs? Can they not be described, depending on the case, as one belief and one mere inclination to believe, or even just as a failure to believe at all? Why assume that it is possible for two so tightly conflicting beliefs to coexist? How is the positive account not just a description of failure to believe, with the label 'belief' attached to it?

To answer this doubt, and develop the positive conception in more detail, let's look briefly at the doubt's most radical form. Schwitzgebel (2010) argues that, in cases like the ones I describe, *neither* of what I call beliefs counts as a genuine belief. Instead, on his view, the examples are of "in-between" believing, an intermediate state that involves no full-blown beliefs at all.

Belief, as Schwitzgebel points out, admits of various intermediate cases. It can be quite unclear whether someone has a belief about an old college dormmate's last name, or believes that God exists, or that all Spanish nouns ending in –a are feminine, or that her son smokes marijuana, or that people of all races are equally intelligent, or that death is not bad (Schwitzgebel 2001, 76-8; 2002, 260-1; 2010, 532). According to Schwitzgebel, on all the leading views of belief, whether dispositional, functional, representational, or other, the dispositions or other roles which constitute belief cover a broad range with intricate variations. Many everyday examples are in-between cases in the gray area between determinately believing and determinately not believing.

In the case of apparently contradictory beliefs, Schwitzgebel's lead example is

"Juliet the implicit racist"(2010, 532 and 543-4). Juliet is someone who

> finds the case for racial equality compelling. She is prepared to argue coherently, sincerely, and vehemently for equality of intelligence and has argued the point repeatedly in the past…. And yet Juliet is systematically racist in most of her spontaneous reactions, her unguarded behavior, and her judgments in particular cases…..To her, the black students never look bright…. When she converses with a custodian or cashier, she expects less wit if the person is black. And so on…. Should we ascribe to Juliet *both* the belief that the races are intellectually equal and the belief that they're not? ...I see little to recommend this approach if it's taken naked: It invites only confusion to say simply…that Juliet both believes that the races are intellectually equal and believes that they are not. For comprehensibility, we need to add qualifications: In such-and-such respects, Juliet acts and reacts as an egalitarian, in such-and-such respects she does not. This is the clearer answer to questions about what Juliet believes; it's also the in-between answer. Does it add anything of value – anything besides confusion – to append to this clear answer the claim that Juliet believes both *P* and its negation? I'm not sure I understand that claim any better than I would understand, in the case of my conditionally reliable computer, a description of it as both reliable and unreliable.

A defender of the possibility of contradictory or akratic beliefs can allow that implicit biases like Juliet's might not be beliefs at all. But Schwitzgebel makes three distinct criticisms that can apply to attributions of both contradictory and akratic beliefs more generally. First, he thinks attribution of contradictory beliefs *underdescribes* the state. Instead of saying "simply" that Juliet believes both, we "need to add qualifications" specifying in what respects she is and is not like an egalitarian. Second, he suggests that attribution of contradictory of beliefs is *empty*. It might not "add anything," on his view, to say someone has both beliefs. Third, he thinks attribution of contradictory beliefs might be *unintelligible*. Schitzgebel doubts that he understands it. These are separate but related objections. For Schwitzgebel, attribution of contradictory beliefs does not tell us enough; once we have the details, it may add nothing at all; and it is hard to even understand what it is saying. One can make the same criticisms of the attribution of akratic belief. What does it add, besides confusion, to say that an anorexic both believes she needs to lose weight, and believes that she shouldn't believe it? Why not just describe the details of her situation, and forgo the dubiously coherent and potentially empty description of her state as akratic belief?

Schwitzgebel's "in-between" view appeals partly to "pragmatic considerations," most centrally its ability to give a "nuanced" view of a person's state (2010, 546; cf. 2002, 270). Other views, he thinks, tend to leave out the nuances. But a positive account of akratic belief can do this just as well. Schwitzgebel's underdescription criticism uses a kind of double standard. To say that Juliet has contradictory or akratic beliefs is not yet to give a full description of her state. But to say *that* she is in an "in-between" state does not

give a full description either. If anything, even on the in-between view, calling someone akratic would at least begin to say more about the way in which she is in-between. But of course it is the further details, concerning *how* she is akratic or in-between, that fill in the picture. Attributing in-between belief does not itself give a fuller description. Nor does it have a clear pragmatic advantage in calling for one. "Somewhere in between" can be a lazy way of ending a discussion; saying that a belief is akratic tends to make us want to know *how* it is.

The suggestion of unintelligibility is also one that Schwitzgebel takes seriously. I have already partly answered it using a series of analogies to other, non-akratic cases in which we recognize belief despite the presence of the same limitations. Another part of the answer comes out more clearly when considering Schwitzgebel's remaining criticism. What does attributing *akrasia* add to his "in-between" description?

First, and perhaps most importantly, attribution of *akrasia* points to the way each belief is integrated into a range of cognitive activity. On the one hand, an anorexic or an avid gambler can decide to change his behavior, find and attend regular support groups, and make a wide range of inferences from his belief that he should not starve himself or should not gamble. He might infer, for example: "Then I probably shouldn't smoke, either," or: "I need to find someone to buy groceries with me," or: "I don't think I can visit my aunt in Las Vegas any more." And he might reason from and act on those inferences in further, complex ways. On the other hand, that same person might often conclude that a course of weight loss or a 2:1 bet on heads is the course of action he should take. And he can engage in more complex reasoning from those conclusions, both in carrying them out and in integrating them with his other commitments, some of which are themselves the results of reasoning from the akratic belief. An akratic anorexic can decide to exercise on a fast day, take a drug to help him get through it, and sneak out of his parents' house at just the right moment to make it work, ignoring the pro-eating reminders he has posted for himself so that they do not lead him to give up his belief that he is fat. What Aristotle (1999, 1142b18) called "calculating" akrasia is no less possible in belief than in action, and is similarly useful in seeing that the case is one of *akrasia*.

Second, attribution of *akrasia* brings out the distinctive role of each belief in inhibiting the other. Without his akratic belief, an anorexic would likely be on the path to recovery. On the other hand, without his belief that he should not believe he should lose weight, his undereating could become less hesitant and more dangerous. Each belief looks partial largely because it has a standing obstacle in the other belief. Unlike attribution of in-between believing, attribution of *akrasia* brings out this peculiarly doxastic tension. It gives us a picture of conflict between beliefs, each of them rationally functioning to a large degree, and each normally ready to manifest fully were it not for the other belief.

Lastly, attributing *akrasia* offers a way to do justice to self-reports of akratic belief. Though it is controversial that belief can be akratic, it is much harder to deny that people can think of themselves as believing akratically. We can self-attribute akratic belief. I've done it myself, and others can too. Someone can say, or think: "I know I shouldn't believe it, but I really do believe bad things happen when black cats walk in front of you." Or: "It

doesn't make sense and I shouldn't even think it, but I'm convinced I need to lose weight."[53] If akratic belief is impossible, reports like these can never be taken at face value. They would have to be indicative of an in-between belief, or some other state, combined with a confusion about the proper attribution of belief in one's own case. If belief can be akratic, there is no need to deny that any such report could be accurate. We can simply allow that an anorexic is self-aware enough to see how distorted his own picture of himself is, even as he continues to act on that picture.[54] For all these reasons, I think we should allow that the component states in "akratic belief" can be beliefs.

Another doubt remains. Even if the 'beliefs' in question are genuine beliefs, are they the *kind* of beliefs that can be akratic? In the case of action, not just anything we believe we should not do will count as akratic if we do it. Akratic action is usually described as voluntary or, in the contemporary literature, intentional. If I fidget, grunt, or scratch my head, my action might be compulsive; perhaps I could not help it; perhaps my movements were not intentional at all. Though "akrasia" is tied etymologically to weakness or lack of self-control, it is, as Owens (2002, 381) puts it, "a failure of control but not an absence of control." Merely compulsive action (if there is such a thing) is not usually thought of as akratic.

If action is only akratic when it is in some way voluntary, intentional, or controlled, akratic belief can seem to require something analogous. This would explain why Rorty "argues…for treating believing as the sort of voluntary condition that can be akratic"(1983, 181). For Mele (1987, 112), akratic belief must "by definition" be "motivated." According to Owens (2002, 388), belief cannot be akratic, partly because "To yield an account of epistemic akrasia,…believing must be purposive; belief must be aimed at a goal." Belief can seem to lack the self-control, and thus the distinctive failure of self-control, distinctive of *akrasia*.

I will argue against such general disanalogies between belief and intention in Chapter 7. If I am right, we have the same kinds of control over intention that we do over belief; so whatever features of intention make it possible for intention to be akratic will carry over to belief as well. But we can already raise a doubt about these views about belief. What sort of feature would make belief appropriately analogous to intention? Must belief be voluntary, motivated, or goal-directed to deserve the label 'akratic'? Examples like the gambler's fallacy, driven by misguided tendencies of reasoning without ulterior motive, should already make us suspicious of such requirements. The interesting questions seem to be: Can we believe what we believe we should not believe? And: can these beliefs manifest a failure of self-control in the realm of belief? To both questions, the answer

[53] These are belief-akratic-paradoxical beliefs or assertions of the kind considered in Chapter 2, Section III.

[54] An analogous point applies to Juliet. In describing Juliet's bias, Schwitzgebel (2010, 532) imagines her "perfectly aware of these facts about herself" and "aspir[ing] to reform." If she is fully aware of the facts, then, on Schwitzgebel's view, she should not straightforwardly see herself as having any particular belief about the intellectual equality of the races. If she insists that she does believe the races are intellectually equal, an "in-between" view would be forced to see her as mistaken about her own beliefs. Why not instead just describe her as conflicted?

seems to be yes. We can recognize, in some cases, a belief that the believer believes she should not have. That belief can be under the believer's control, in some of the ways beliefs normally are. It can be based on apparent evidence, inferred from other beliefs she holds, and incorporated into a larger chain of thought that gives rise to action. At the same time, the belief shows a kind of failure of self-control, since the believer is unable to bring it in line with what she believes she ought to believe. An akratic anorexic can defy her better judgment by believing she is fat, believing she needs to lose weight, and orchestrating an elaborate and painful series of weight loss regimens even while, convinced by her doctor, she understands that she is malnourished and should not believe she is fat or in need of weight loss. We see here the details of one kind of striking inner conflict in which we sometimes find ourselves—so much so that it can begin to seem strange that akratic belief ever seemed puzzling or impossible. I turn next to explaining the puzzlement.

IV. Why is Akratic Belief Puzzling?

If I am right, the examples of akratic belief are legion. Anorexia, fear of flying, the gambler's fallacy, skepticism, and belief in a friend's trustworthiness are just a handful of recognizable contexts. Just about anything one can believe, it seems, is something one can simultaneously believe one should not believe. On the other hand, the very possibility of believing akratically seems puzzling—and, I think, rightly so. But if akratic belief is possible, why does it seem so strange? Can we explain what makes akratic belief puzzling, while still maintaining its possibility?

Akratic belief is puzzling, partly in the ways akratic action is puzzling. One wants to ask: If you think you shouldn't be doing it, why are you doing it? If you think you shouldn't believe it, why do you? The answers given in particular cases can seem at best unsatisfying. Someone might say: "Because eight tails in a row *never* happens." In a case of akratic acquiescence to the gambler's fallacy, this reasoning is bad, even from the speaker's point of view. Eight tails in a row is a very rare outcome; but for fair coins that just saw seven tails in a row, it happens 50% of the time. This is not just obvious; it is accepted by the speaker. Bafflement is a natural reaction. If the answer is instead: "No reason," or: "I just do believe it," a similar bafflement is natural. Our ordinary practices of asking for justification are constantly frustrated by akratic cases. This is part of why akratics are so hard to interact with. And of course, the puzzlement arises not only in interacting with *akrasia*, but in the mere contemplation of it. It is puzzling that someone can be so disunified, and still be a single person.

But there is more to puzzlement about akratic belief. One can ask: why does akratic belief more commonly seem impossible than akratic action does? Akratic belief can seem impossible even to imagine, both in general and in particular cases. In one's own case, for example, it rarely takes long to remember at least one akratic action. Most of us akratically overeat, or snap at someone, or stay up late, or go to bed without flossing, if nothing else.

But it can be hard to think of a single example of one's own akratic belief. Is there some explanation for this? Can we say why philosophers still make general denials of the possibility of akratic belief, without denying the possibility of akratic action? The sense that akratic action is puzzling cannot explain all of our puzzlement about akratic belief.

Relatedly, one can feel that it is in some way 'harder' to maintain an akratic belief than it is to complete an akratic action. It seems so easy for akratic belief to collapse into giving up one belief or the other, or into "in-between" belief, or into simple uncertainty or suspension of belief. This is a thought that a description of akratic belief should be able to address. Even if akratic belief is possible, it can seem to be an especially unstable state, and in that sense less easy to find oneself in or imagine. Is there something right about this appearance? Is akratic action in some sense easier?

These kinds of puzzlement have more than one source. Explaining them fully might be impossible without appeal to a particular theoretical conception of belief and action. Even so, there are at least four partial explanations of the sense that akratic belief is impossible or especially 'difficult' to maintain.

First, prior commitment to a theory about belief, intentional action, or *akrasia* can itself make akratic belief seem more puzzling. If we assume that a person's beliefs must be consistent, or deny that beliefs exist, akratic belief is likely to seem impossible. Or if we believe, with Rorty, Mele, or Owens, that *akrasia* in general must be voluntary, motivated, or goal-directed, belief can seem unable to count as akratic. We can then lose sight of the possibilities of holding, and reasoning with, beliefs that we ourselves believe we should not have.

Second, belief, however we understand it, is normally thought of as an ongoing state. In this respect, belief resembles intention rather than action. So the natural analogue to akratic belief might be not akratic action, but akratic intention: an intention one believes one should not have. Akratic intention can already seem less familiar, and more puzzling, than akratic action is. Some of us may be committed to a theory on which it is impossible, or more difficult, for two states—an intention and a belief—to stand in akratic conflict with each other than it is for an action to conflict with a belief. But even without a theory, the 'ongoing' character of intention can be a source of puzzlement about the possibility of akratic intention. Though actions have temporal duration, it is natural to think of intention as lasting much longer than the corresponding action does. It can be especially puzzling that I would intend all year to diet next summer, believing all the while that I should not intend it. How can I go all year without reconsidering the intention? Or to take a more standard example, I can take an extra scoop of ice cream after dinner against my better judgment; but how can I intend all day to take it? In the case of an ongoing intention that precedes action, there is often a less powerful temptation to go against one's own belief, and, at the same time, a longer opportunity to give the intention up. Akratically intending all year to diet, or all day to overeat, can thus seem more puzzling than their corresponding actions, and in one sense harder. One must do more to avoid or resist the implications of one's own beliefs. If intention, rather than action, is the practical analogue of belief, one should expect that akratic belief would seem more puzzling or harder to maintain than

akratic action, just as akratic intention does.

Third, there might be a real difference with a contingent psychological explanation. Akratic action and intention might in fact be more common than akratic belief in humans as a biological species. Pears (1982, 50) writes: "Vividness and other similar qualities of perceptual cues have much less force than the special qualities of physical appetites which make them such successful rebels." Comparisons of the force of appetites and perceptual cues might have to be imprecise at best. But in explaining an intuitive sense of the difficulty of akratic belief, the thought is useful. Those of us who are especially carried away by the vividness of perceptual cues, the misleading appeal of fallacious reasoning, or, in some cases, wishful thinking or self-deception, might be especially prone to akratic belief. The rest of us might be mostly immune, just as the most virtuous or continent among us rarely or never act akratically. More generally, the difference in force is likely to be at least partly contingent on the details of human psychology. Humans might be stronger—more able to resist temptation—in belief than in action, whereas other species, real or imagined, might be mostly continent in practice but terribly susceptible to the gambler's fallacy, even when they know better. Relative rarity and difficulty are not always signs of an underlying metaphysical difference.

This third explanation can seem essentially misguided. After all, desire and appetite can affect belief as well as action. Even if our species is not as prone to be misled by perceptual cues as we are by motivational ones, can our desires not give rise to akratic belief through self-deception or wishful thinking? And if they can, what would the contingent psychological difference be?

Pointing to the possibility of motivated akratic belief does not obviously threaten the possibility of akratic belief. But it does raise a challenge for an explanation of any distinctive puzzlement about akratic belief. If both akratic belief and akratic action can arise through the influence of desire, why is the belief more puzzling? I think the answer is threefold. First, although akratic wishful thinking and self-deception are not easily shown to be more rare than akratic action, they might still not usually result in an akratic belief. We are often left deceived about what we ought to believe as well as in the first-order belief itself. Akratic belief might still be less common than either akratic action or wishful or self-deceived belief. Second, even when we form an akratic belief through self-deception or wishful thinking, we might not notice that we do it. At least often, self-deception and wishful thinking depend on a lack of awareness about their own operation. If they sometimes stop us from noticing our own akratic belief, akratic belief can seem even less common than it actually is. Third, wishful thinking and self-deception are themselves notoriously puzzling. Many of us find them strange, and wonder how they are possible. So rather than calling into question the appeal to a contingent difference between *akrasia* in belief and action, the possibility of wishful thinking and self-deception only complicates our understanding of that difference, and adds a further source of puzzlement about akratic belief. Akratic beliefs maintained by wishful thinking or self-deception can seem strange, partly because wishful thinking and self-deception seem so strange.

It is worth mentioning a fourth and final partial explanation of the distinctively

puzzling quality of akratic belief. The sense that akratic belief is hard to find, imagine, or maintain might itself be partly the product of the limited imagination of those of us who rarely find (or recognize) ourselves to be believing akratically. Akratic actors can feel a tinge of intellectual superiority over those who, used to acting as they believe they should, profess themselves incapable of understanding how anyone could ever fail to follow their own better judgment. For someone in a constant struggle with anorexia, superstition, or the gambler's fallacy, denials of the possibility of akratic belief can themselves show a surprising lack of imagination. For her, akratic belief can be all too easy to find, imagine, and maintain; the difficulty can instead be in reaching, or in some cases even imagining, a state in which one's beliefs are what one believes they should be. Those of us who are relatively immune to akratic belief, or relatively bad at recognizing it in ourselves, can underestimate how stable akratic belief can be.[55]

By itself, this last explanation might not point to a difference between akratic action and belief. Our imagination might be just as limited in both cases. But a difference does emerge when the third and fourth explanations I offered are combined. If akratic belief is less common than akratic action, it can be especially difficult to imagine. In the case of akratic action, our limited imagination is not needed; we have examples of akratic action all around us and, all too often, in ourselves. But in the case of belief, more of us do need some imagination. Akratic belief might indeed be less common in our species than akratic action. And its distinctive kind of conflict makes it difficult to imagine from the point of view of the akratic believer herself. It can thus take some reminders and some careful description of a range of examples to restore insight into the plight of an akratic believer. The situation of an anorexic, or gambler, or superstitious person with an akratic belief is, for many of us, both foreign and complex. Akratic belief can indeed be more puzzling, and less common, than akratic action; but it is not therefore completely inconceivable.

I have tried to address our natural puzzlement about the possibility of akratic belief. I do not mean to dissolve the puzzlement. We *should* be puzzled. Akratic belief shows a worrying and intensely conflicting mismatch between our own beliefs and what we ourselves believe they should be. It is a striking failure of cognitive integration. Puzzlement is appropriate. One mark of an understanding of *akrasia* is its ability to explain, not only how *akrasia* is possible, but why it is puzzling.

Still, if I am right, the impossibility of akratic belief is a philosopher's fiction. Though it can be an implication of a theoretical conception of belief, it should count against that conception. We can recognize akratic belief in ways similar to the ways we recognize any other belief, albeit with some more difficulty and bafflement. And if a view about the nature of belief does not recognize akratic belief, we should be more hesitant to

[55] For another, more detailed discussion of the relative 'ease' of akratic action, see the rest of Pears (1982). Since Pears does proceed by considering particular theories of action, I leave out the details here. But it is interesting to note that Pears combines, and perhaps confuses, intuitions of impossibility and of difficulty when he writes that motivated, "full-blown" akratic belief is "scarcely" or "only marginally possible"(44,46,49,50).

recognize the theory as true. A theory that rules out the possibility of akratic belief is, for that reason, less believable.

We can explain both how belief can be akratic, and why akratic belief is puzzling. We can also explain why it can seem puzzling that people find it so puzzling. There is something odd about the expectation of such coherence in an ordinary human life—something out of touch with the striking divisions within a single person's patterns of reasoning and conviction. Like the denial of the possibility of akratic action, puzzlement about the possibility of akratic belief has an air of blindness to the conditions of life, at least when that puzzlement reaches the point of denying the possibility of akratic belief.

The point can be put less critically. Accepting the possibility and the variety of akratic belief is part of having a lifelike picture of ordinary cognition. It is useful for compassionate and resolute interaction with those who are especially prone to *akrasia*. It might lower a natural resistance to recognizing it in one's own case. It shows us some of the limits of the thought that each of us has a single, unified point of view. And, I think, it prevents us from drawing a misleading disanalogy between theoretical and practical reasoning. In both, the conclusions we believe we should reach can differ starkly from the ones we actually come to.

The possibility of akratic belief has at least one other theoretical implication. Theories that entail, or face pressure to accept, the possibility of akratic belief are at an advantage, not a disadvantage. This implication brings us back to the defense of the Identity View, with which Chapter 2 began. As I will argue in the next few chapters, the view that intention is a normative belief can be defended against challenges to it, in ways that illuminate the details of ordinary activity. The apparently problematic implication that akratic belief is possible is one example. I turn next to akratic action, which, on the Identity View, can itself seem puzzling and even impossible.

Chapter 4: Acting Against One's Better Judgment

It is hard to deny that we often do what we believe we should not. We believe, for example, that we should not eat dessert, or stay on the computer, or insult our friend. But we do it anyway. These actions are often called akratic, or sometimes weak-willed, or "incontinent," or against one's own better judgment. For A to *x* akratically is for A to *x* intentionally, while believing she should not *x*. For A to akratically *intend* to x, we can say, is for A to intend to x, while believing she should not x.[56] This is an immediately recognizable and puzzling phenomenon. It is especially puzzling for a guise-of-the-good view of action or intention.

Recall

The Identity View: A's intention to x *is* a belief that A ought to x.

On this view, someone who intends to eat dessert must believe she ought to eat it. The Identity View requires a kind of normative endorsement, not disapproval or prohibition, in every case of intention. If we act and intend akratically, how can any such view be true? How can evaluation and motivation line up so neatly if they so obviously come apart? In this chapter, I offer a defense of the Identity View on this central topic, by explaining how it can allow for akratic action and intention. The purely defensive aim is to show that akratic action and intention provide no compelling counterexamples to the Identity View. But I will also try to show how the Identity View can shed light on the details of *akrasia*.

I will start by trying to get clear about the problem. In §I, I argue that *akrasia* presents an important challenge to the Identity View, but one that is different and less intractable than the one it seems to present. The challenge is to show that when we have an akratic intention, we believe we ought not do something, and *also* believe that we ought to do it. The explanation of *akrasia* lies in conflicting normative beliefs. Drawing on the discussion of belief in Chapters 2 and 3, I will argue in §I that we should allow on independent grounds that conflicting normative beliefs are possible in principle. In §§II-III, I argue that there is no compelling obstacle to attributing such beliefs in *all* cases of akratic intention. In §IV, I argue that we can see these cases as cases of conflicting intentions as well as conflicting beliefs. As a whole, this chapter argues that the Identity View is both defensible and illuminating with respect to akratic action and intention. It is conflict

[56] I continue to use 'ought' and 'should' interchangeably, and from here on I will focus mostly on intention, rather than directly on intentional action. I leave aside the issue of whether intentional actions must be actions we intend to perform. If we can do something intentionally without intending to do it, such actions present no further problem for the Identity View.

between normative beliefs, rather than a simple mismatch between intention and belief, that underlies the phenomenon of *akrasia*.

I. The Possibility of Conflicting Beliefs

It seems clear that we can believe we should not do something, and still do it anyway. How can anyone deny that?

The key initial point is that the Identity View makes no such denial. It does not hold that intentional action or intention precludes negative evaluation. Instead, it attributes a *positive* evaluation—a belief that we ought to act as we do. We should distinguish *lacking* a normative belief, from *having* a 'negative' one, or believing in a prohibition. The Identity View requires a belief that we ought to act as we intend to. When we act akratically, we believe we *ought not* act as we do. That is different than *not* believing we ought to act as we do. So strictly speaking, there is no immediate problem about believing we ought not do something and doing it anyway. There is always the possibility of believing that one ought to *and* believing that one ought not. Akratic intention is possible when we have conflicting normative beliefs.

To bring this out, consider Donald Davidson's classic account of *akrasia*. In "How is Weakness of the Will Possible?", Davidson describes the problem as a conflict between three principles that all "seem self-evident"(1980a, 23):

> P1. If an agent wants to do x more than y and believes himself free to do either, he will intentionally do x if he does either intentionally.
> P2. If an agent judges it better to do x than y, he wants to do x more than to do y.
> P3. There are incontinent actions.

P3 simply restates what it seems we already knew: that it is possible to act intentionally against one's own better judgment. In Davidson's (1980a, 22) terms,

> In doing x an agent acts incontinently if and only if: (a) the agent does x intentionally; (b) the agent believes there is an alternative action y open to him; and (c) the agent judges that, all things considered, it would be better to do y than to do x.[57]

[57] One complication is that Davidson, and others following him, define *akrasia* in terms of available alternatives. As Tenenbaum (2007, 257) puts it, "An akratic agent will think that A is better than B yet pursue B." Here I temporarily go along with Davidson's formulation, though I think a non-comparative one is both simpler and more precise. You may, for example, believe you should not smoke, but not have thought about how else you would spend your next five minutes. If someone does not have a particular better alternative in mind, but believes she should not do B, her doing (or 'pursuing') B is still akratic in the central sense of being against her better judgment, or something she believes she ought not do.

P3 simply tells us that we do sometimes act this way. And yet the existence of such actions seems to fly in the face of "another doctrine that has an air of self-evidence: that, in so far as a person acts intentionally he acts…in the light of some imagined good"(1980a, 22). P1 and P2 together reflect that other doctrine. Davidson thinks the conflict cannot be resolved: "No amount of tinkering with P1-P3 will remove the underlying problem"(24). He then goes on to develop a conception of *akrasia* that attempts to preserve all three principles.

P2, which connects judgment with wanting, is the principle that most directly expresses a guise-of-the-good view. It has been common to resist guise-of-the-good views by rejecting P2.[58] But I think this rejection is a misguided way of resisting a guise-of-the-good view. It is misguided, because a guise-of-the-good view should *itself* reject P2. P2 says: "If an agent judges it better to do x than y, he wants to do x more than to do y." By the same token, if an agent judges it better to do y than x, she wants to do y more than to do x. So if her judgments conflict—if she judges both that it is better to do x than to do y, and that it is better to do y than to do x—then she must both want to do x more than to do y, and want to do y more than to do x. That is not just irrational; it is impossible. Two desires cannot each be stronger than the other. So if P2 is right, it is impossible to both judge it better to do x than y, and judge the contrary.[59]

Davidson does not explicitly consider whether simply not doing something can count as an action, though in a reply to Bruce Vermazen (1985b, 217), he allows that it can. If simply refraining—not doing B, not smoking—counts as an alternative action, the comparative definition and the non-comparative one might apply to at least close to the same cases. But there would still be two differences. First, the comparative definition has the person thinking in terms of what is better rather than what she ought to do. Second, even without that difference, the comparative definition has the person believing she should instead refrain, rather than, more directly, that she should not do B. These differences can be distracting. Rather than asking how significant the differences are, I use the simpler formulation: we act akratically when we intentionally do what we believe we ought not do.

[58] This line of resistance has been taken in some now classic treatments of akratic action; see Watson (1977), Audi (1979), Pears (1982), and Mele (1983).

[59] Can we revise P2 to avoid this conclusion? Davidson writes that "a problem about incontinence will occur in some form as long as there is any word or phrase we can convincingly substitute for 'wants' in both P1 and P2"(1980a, 27). But the problem about conflicting judgments itself applies to any version of P1-P2 in which the same phrase is substituted for "wants"—that is, any version in which the two principles are linked by a common phrase. We can see this by removing the phrase. The combined form of P1-P2 would be:

> P12. If an agent judges it better to do x than y, and believes himself free to do either, he will intentionally do x if he does either intentionally.

By the same token, if an agent judges it better to do y than x, and believes himself free to do either, he will intentionally do y if he does either intentionally. But now imagine someone with conflicting judgments: someone who both judges it better to do x than y, and judges it better to do y than x. According to P12, if this person believes himself free to do either, he will do both x and y if he does either intentionally. In akratic cases, x and y must be mutually exclusive, since otherwise it could easy to judge x better than y and still do y, by doing both. Doing both x and y is impossible,

On the Identity View, an intention to eat dessert is a belief that one ought to eat dessert. To akratically intend to eat dessert, as I have described it, is to believe one ought to eat dessert, while believing one ought not eat it. Principles like Davidson's may lead us to think that such conflicts are impossible. In that case, akratic intention would be impossible on the Identity View. But I think we should allow that we are capable of conflicting normative beliefs. We can allow this on independent grounds, which do not assume any guise-of-the-good view to be true. And allowing the possibility of conflict is a help, not a hindrance, to a guise-of-the-good view. The challenge for the Identity View is not to prove the impossible. It is not to show that we can believe we ought to do something that we do not believe we ought to do. Instead, it is to defend a conception of akratic intention as a particular kind of conflict. This is what I do in this and the next two sections: first, by defending the view that conflicting normative beliefs are possible in principle; second, by defending the attribution of them in cases of *akrasia*; and third, by developing a parallel conception of conflict in intention.

To avoid ruling out the possibility of conflicting beliefs, we can formulate

> *The Conflict Constraint*. An accurate conception of *akrasia* must allow that a person can both believe she ought to do something, and believe she ought not do it.

The Conflict Constraint insists that we allow the possibility of conflicting normative beliefs: beliefs that together require both performing an action, and not performing it. I call it "The Conflict Constraint," rather than "The Contradiction Constraint," because what it requires a conception of *akrasia* to allow is not contradictory beliefs, of the form "p" and "not-p", but normative conflict, of the form "I ought to x" and "I ought not x." Such a conflict both requires and forbids an action, rather than requiring the action and denying the requirement.

There are two available senses of 'conflict in belief' here: conflict *within* a belief, and conflict *between* beliefs. A single belief can show normative conflict, if the belief is of the form: "I ought to x and I ought not x." For normative conflict *between* beliefs, it is enough to have a belief of the form: "I ought to x," and a belief of the form "I ought not x." These beliefs are in normative conflict with each other, even without a belief in the conjunction: "I ought to x and I ought not x."

The Conflict Constraint requires only that a conception of *akrasia* allow the possibility of conflict *between* beliefs. But distinguishing these two kinds of normative conflict also allows us to distinguish two ways of coming to accept the Conflict Constraint.

One begins with examples of apparent within-belief conflict. A sharp conflict of

---

judging as wanting to do each more than to do the other is impossible. So if P12 is true, it is impossible to have conflicting judgments about which one is better, believe oneself free to do either, and do either of them intentionally. This conclusion is quite general. It says that in any situation in which we are presented with alternative actions, it is impossible to have conflicting judgments about which one it is better to pursue and still intentionally pursue one. As I go on to argue, guise-of-the-good views should reject this conclusion.

this kind can be seen in Thomas Nagel's essay "War and Massacre." Nagel describes some especially stark moral dilemmas related to war, such as the question of whether to torture a terrorist. He suggests that situations like these can put us into what he calls a "moral blind alley," in which, for example, torturing the terrorist would be wrong, and not torturing the terrorist would be wrong. He writes (1972, 143-4):

> The idea of a moral blind alley is a perfectly intelligible one. It is possible to get into such a situation by one's own fault, and people do it all the time. If, for example, one makes two incompatible promises or commitments—becomes engaged to two people, for example—then there is no course one can take which is not wrong, for one must break one's promise to at least one of them. Making a clean breast of the whole thing will not be enough to remove one's reprehensibility. The existence of such cases is not morally disturbing, however, because we feel that the situation was not unavoidable: one had to do something wrong in the first place to get into it. But what if the world itself, or someone else's actions, could face a previously innocent person with a choice between morally abominable courses of action, and leave him no way to escape with his honor? Our intuitions rebel at the idea, for we feel that the constructibility of such a case must show a contradiction in our moral views. But it is not in itself a contradiction to say that someone can do X or not do X, and that for him to take either course would be wrong. It merely contradicts the supposition that *ought* implies *can*—since presumably one ought to refrain from what is wrong, and in such a case it is impossible to do so.

I quote this passage not to agree or disagree with it, but simply to point out that Nagel himself seems to believe it. He suggests that it can be true that doing X or not doing X can both be wrong, in the sense that one ought not do X and ought not *not* do X. And so, it seems, Nagel believes that in some situations, for some X, we ought to X and ought not X. About particular situations, he may have beliefs with a content like: "I ought to torture the terrorist and I ought not torture him." Since Nagel seems to have in mind a genuinely normative 'ought', rather than a specifically moral but normatively escapable one, we can talk directly of "normative blind alleys." To endorse the possibility of a normative blind alley is to endorse the possibility that a belief with a within-belief normative conflict can be true. To *believe oneself to be in* a normative blind alley in a particular case is to have a within-belief normative conflict. Although we may not entirely understand what it is like to have such a conflict, seeing someone insist, and argue, that she is in one may lead us to admit that it is possible.

The possibility of within-belief conflict, it is natural to think, entails the possibility of between-belief conflict. Most of us think that belief distributes over conjuncts: that someone who believes "p and q" already believes "p" and believes "q." And even if belief does not distribute over conjuncts, someone is normally able to infer each conjunct from the conjunction. Someone who believes herself to be in a normative blind alley believes in,

or at any rate can infer, each of its components. For someone who believes "I should torture this terrorist and I should not torture him," both "I should torture this terrorist" and "I should not torture him" seem to be, at the very least, easy conclusions to draw from the conjunctive belief. According to this line of thought, within-belief conflict is possible; if within-belief conflict is possible, between-belief conflict is possible; so between-belief conflict is possible. We can call this the *Argument from Within-Belief Conflict*.

The inference in the other direction can be more difficult. An anorexic who believes "I should lose weight" and believes "I should not lose weight" might find it far from trivial to form the conjunctive belief "I should lose weight and I should not lose weight." Within-belief normative conflict can be harder to maintain than between-belief conflict, and on a more restrictive view, it may not be possible. Some may think that the notion of a normative blind alley makes so little sense that Nagel himself cannot understand it well enough to ever believe himself to be in one.

Fortunately, there is no need to rely on the possibility of within-belief conflict. We can, instead, attribute between-belief conflict directly. In Chapter 3, I described some typical marks by which we attribute belief: we often look for sensitivity to evidence, or to apparent evidence; recall in relevant circumstances; felt conviction; reporting of the belief to others; and use in further reasoning. In, for example, conflict about weight loss, each of two beliefs can manifest these characteristic features. Each can be sensitive to apparent evidence: one to an apparent horrible overabundance of fat seen in the mirror, the other to a doctor's warnings of malnutrition. Each can be reported sincerely in various contexts and to different people, recalled often and with conviction at various times, and so on. And it is not always easy to settle on a single belief attribution statistically, based on a greater frequency or duration of dominance of one belief over another. Someone's thoughts and behavior can be so starkly contradictory that it can be described only by attributing two conflicting normative beliefs to him. This can be true even though it can be difficult for him to combine these beliefs into a single perspective. "I should lose weight and I should not lose weight" is much harder to support with evidence, to hold with any conviction, to sincerely report, and so on. Between-belief conflict can be easier to maintain, and more resistant to resolution. So although accepting the possibility of within-belief conflict is one way to accept the possibility of between-belief conflict, it is not the only way. The usual characteristics by which we recognize belief can themselves conflict, and can call for the attribution of two conflicting beliefs. We can call this the *Argument from Belief Attribution*.

To see the force of these arguments for the possibility of conflicting normative beliefs, it is helpful to compare the case of believing straightforwardly contradictory propositions. According to Davidson, we can believe two contradictory propositions, but not their conjunction. As he put it (1985a, 198): "It is between these cases that I would draw the line: someone can believe *p* and at the same time believe not-*p*; he cannot believe (*p* and not-*p*)."[60] For example, it might be possible to believe that it is raining, and believe

[60] See also Davidson (1986) and (1997).

that it is not raining. But, for Davidson, I could never believe that (it is raining and not raining). Here he sets a limit on the possible extent of incoherence.

There is something to be said for this view. When we have a belief, we tend to be sensitive to the evidence for it; recall it in relevant contexts; reason from it; feel some conviction in it; and report it to others. We might be able to do this for two contradictory propositions; each belief might be sensitive to some of the available evidence, felt with at least intermittent conviction, and recalled, reported, and reasoned from at least some of the time. This is not as obviously true of beliefs of the form: "*P* and not-*p*." [61] It is natural to think that there is no evidence for such beliefs; no particular relevant circumstances in which to recall them; and, perhaps, no way to feel conviction in them, form further beliefs on the basis of them, or sincerely report them to others. Even if two contradictory beliefs, taken separately, can maintain their character as beliefs, it is natural to think that belief in the corresponding conjunction cannot. If Davidson is right, an anorexic cannot believe: "I need to lose weight and I do not need to lose weight." But he can still believe that he needs to lose weight, and believe that he does not.

Davidson's view seems attractive, partly because it is hard to imagine how someone could believe a proposition of the form "*P* and not-*p*." But it is conceivable that there could be examples that would make this easier to imagine. One such example is provided by dialetheists, who believe that some statements are both true and false. Graham Priest writes (2006, 96-7):

> There are many cases where people consciously believe an explicit contradiction (and with no real doubt)….I, for example, believe that the Russell set is both a member of itself and not a member of itself. I do not deny that it was difficult to convince myself of this, that is, to get myself to believe it. It seemed, after all, so unlikely. But many arguments, most of which appear in this book, convinced me of it.

As with Nagel, the key point is simply that Priest seems to believe what he says he believes. Priest is reporting the belief that the Russell set is both a member of itself and not a member of itself. As he says here, he appeals at length to apparent evidence for this belief. He also recalls it here as relevant, seems to feel conviction in it, and appears ready to reason from it. Nor does his belief seem fleeting or unstable. It probably still seems puzzling. One can wonder what it is like to believe such a thing. But even if his book does not convince us that the Russell set both is and is not a member of itself, the book may convince us that he believes it. As puzzling as the belief is, Priest does seem to have it. And if he does, surely he either already believes, or at least can easily come to believe, that the Russell set is a member of itself, and at the same time that it is not. This is the analogue

[61] There are other ways to motivate this denial. For example, proponents of truth-conditional semantics can motivate it in a semantic way: To understand a statement is to understand under what conditions it is true, and contradictions, many of us think, are not true under any conditions. They may then not even make sense, let alone be believed.

of the Argument from Within-Belief Conflict.

As before, accepting the possibility of explicit or conjunctive conflict is not essential. If such contradiction is too incoherent to be believed, Priest must be misattributing the belief to himself. We can ask what the grounds are for thinking he must be making a mistake in attribution. But even if he is, we can still make out a conflict between his beliefs. For each component belief—that the Russell set is a member of itself, and that it is not—we can see ample conviction, recall, apparent evidence, and further reasoning in Priest's writing. A self-proclaimed dialetheist will even tend not to waver in the signs of one belief when confronted with the evidence for the other. All the characteristic marks of belief seem standardly present in each case. This is the analogue of the Argument from Belief Attribution.

David Velleman doubts that a guise-of-the-good view can articulate a necessary characteristic of intentional action or intention. For him, such a view describes not agency in general, but "a particular species of agent, and a particularly bland species of agent, at that."(2000, 99). I think it is the denial of the possibility of conflicting normative beliefs that describes "a particularly bland species of agent," and insists on a naïve optimism about the coherence of our attitudes. There is no reason to insist that this and other forms of conflict are impossible. Both everyday life and philosophical theory provide many examples of strikingly extreme conflict and inconsistency. A guise-of-the-good view can accept that there are such conflicts. If I am right, the Identity View depends on their possibility. So from here on, I will assume that conflict between normative beliefs is possible, and that the Conflict Constraint is true.[62]

The problem of akratic action is: how can one believe that one ought not do something, and still intend to do it? In one way, this problem has an easy answer, even if we accept the Identity View. We intend to do it, and we can intentionally do it—take an extra helping of ice cream, for example—because we do believe we ought to, and we act on that belief. Believing we ought to do something does not stop us from also believing we

[62] Davidson is not the only one to ignore the possibility of such conflict. Mele (1983, 357-8), for example, argues that since we can intend against our own better judgment, intention cannot itself be such a judgment—a conclusion that follows only if we rule out the possibility of conflicting judgments. Bratman (1979, 157) gives an example of Sam, who drinks after deciding it would be best not to. As Bratman sees him, "Sam surely does not also conclude that it would be best to drink; though guilty of some form of irrationality, Sam is not guilty of such blatant inconsistency"(1979, 157). Bratman then adds, in a somewhat different context: "I assume that the agent does not hold logically inconsistent views" (1979, 171n13). On the view I am developing, such conflict—though perhaps not properly called "logical"—is central to akratic action, and its possibility undermines any theory that depends on ignoring it.

Even defenders of guise-of-the-good views sometimes ignore the possibility of conflicting beliefs or judgments. Tenenbaum writes (2007, 14): "An agent desiring X is to be identified with X *appearing* to be good to the agent…, not with the agent *judging* it to be good. This small shift guarantees that the view does not fall prey to the most obvious objections to it; the scholastic view, for instance, does not deny that we can desire what we know is not good." If we can know something is not good, *and* judge it to be good, there is no obvious objection here.

ought not. Unfortunately, believing we ought not does not stop us from believing we ought to, either. We can act akratically, because our beliefs can conflict in this way.

The basic idea, once again, is that *akrasia* itself involves belief that conflicts with our 'better judgment'. Reaching for dessert has an evaluative structure.[63] Typically, the dessert suddenly strikes us as delicious and as something we ought to have, even if we disapprove of our own motivation and believe we ought to skip dessert. The possibility of conflicting beliefs takes some of the bite out of the counterexample. It prevents *akrasia* from providing a direct refutation of the Identity View.

The possibility of conflict makes accounting for *akrasia* difficult instead of impossible. But *akrasia* still presents a challenge. The Identity View can allow that we can intend to do what we believe we ought not do. But it does deny that we can intend to do what we do *not* believe we *ought* to. It then must hold that *whenever* we intend to do something we believe we ought not do, we must *also* believe that we ought. It has to hold, in other words, that *akrasia* always involves conflicting normative beliefs. My next challenge is to defend this view.

II. The Conflicting Belief View

We can call the view I am defending

The Conflicting Belief View. Akratic intention requires conflicting normative beliefs.

Though this view is rarely taken seriously, I believe it is defensible. In this section, I develop the view, beginning with some historical precedent.

In Book VII of his *Nicomachean Ethics*, Aristotle presents an account of *akrasia* and uses it to address a series of puzzles. One of them is about the akratic person's knowledge. Against the background of Socrates' denial of 'clear-eyed' or knowing *akrasia*, Aristotle insists that there is such a thing as knowingly acting contrary to one's own judgment about what to do. He then tries to explain in what way the akratic person knows that what she is doing is wrong.[64]

---

[63] Aristotle famously makes a brief mention of the possibility that "The incontinent or base person will use rational calculation"(1999, 1143b18; cf. 1149b14-18). The view I propose is not that all akratic action is calculating; it leaves open the possibility that an akratic action might involve normative belief without any means-end reasoning or other calculation. But since acting on an evaluative belief usually does involve some inference about means, I am, in effect, treating calculating *akrasia* as the paradigm case.

[64] As Broadie (1991, Chapter 5) emphasizes, Aristotle's task is not to explain how *akrasia* is possible, or to give a full empirical account of its workings, but, most centrally, to account for the akratic's knowledge. I do not think Aristotle is right to speak of knowledge rather than belief, since acting against one's better judgment is naturally seen as akratic even when the better judgment is

This is a difficult task. As Aristotle sees, knowing what one ought to do is closely connected to actually doing it. In someone who knows that an action is wrong, and attends to this knowledge, doing the wrong action "seems extraordinary"(1999, 1146b36). In discussing practical reasoning in cases of production, Aristotle says that "it is necessary…to act at once on what has been concluded"(1999, 1147a27-29). In general, without the interference of appetites, he thinks, practical reasoning leads to action. So in *akrasia*, where one does not do what one knows is right, something must be going wrong in deliberation. We can then ask: (1) What goes wrong, preventing the reasoning from giving rise to action? (2) What goes right, allowing the akratic person to still count as having knowledge?

Nicomachean Ethics VII.3 tries to answer this pair of difficult questions. Aristotle introduces two key parts of his answer early on when he says that the akratic person "uses only the universal premise"(1999, 1147a2-3). Apart from locating the trouble within deliberation as he understands it, Aristotle is also drawing a distinction between having knowledge and using it. That distinction is a helpful one, and easily recognized. Everyone has knowledge that they do not attend to or act on, without forgetting it. One normally knows one's address, no matter where one is. But if we are far away and focused on something else, we may have this knowledge without 'using' it.

Aristotle also suggests a distinction between ways or degrees of having knowledge:

> Having without using includes different types of having; hence some people, such as those asleep or mad or drunk, both have knowledge in a way and do not have it. Moreover, this is the condition of those affected by strong feelings.
>
> 1999, 1147a11-15

> The last premise…is what the akratic does not have when he is being affected. Or [rather] the way he has it is not knowledge of it, but…[merely] saying the words, as the drunk says the words of Empedocles.
>
> 1999, 1147b10-14

> And those who have just learned something do not yet know it, though they string the words together; for it must grow into them, and this takes time. And so we must suppose that those who are acting akratically also say the words in the way that actors do.
>
> 1999, 1147a21-24

Here Aristotle is describing various ways of having and not fully having knowledge, and placing *akrasia* in the category of less than full knowledge.

His analogies are often seen as unhelpful and even frustrating. They make a series of quick comparisons to a wide variety of phenomena that seem to have little in common:

---

mistaken, and therefore not knowledge. But it is not hard to see how the same line of thought can be taken when thinking of belief rather than knowledge.

sleep, insanity, drunkenness, strong feelings, learning, and theater. Nor do most of these seem to have much in common with *akrasia*. Even the descriptions can seem off; "merely saying the words" and "string[ing] the words together" suggest empty lip service, which is not what the akratic person does when she expresses her judgment.

But Aristotle has a point, and these remarks are much more illuminating than they might sound. The wide range of the comparisons is precisely chosen, and helpful for Aristotle's explanation. It brings out the key characteristic that all these various things share, and that Aristotle means to draw attention to. They all involve a kind of impaired knowledge or belief whose expression is more than empty lip service. As Sarah Broadie (1991, 296) puts it, Aristotle "is not claiming that the agent can utter the words as meaningless noises (what would be the point of that claim?) but that the saying does not express what it should, i.e., the actively serious purpose which is the grasp of practical truth." On the other hand, the comparisons also suggest that the saying does express something. As Broadie later says, an actor does *feel* his words. Actors enter imaginatively into what they say, and even have a kind of commitment to it which can lead them to further action. But the commitment is not complete and, importantly, not stable. Similarly, a drunk who says "This is my eighth drink" has some awareness of the fact, and even of its significance. But his awareness is hazy, and, like the actor's, unreliable and limited.

All of these examples involve knowledge, or belief, that is impaired in a particular way. (Aristotle himself does not see his focus on knowledge as essential; "whether it is knowledge or belief," he thinks, "does not matter for this argument"(1999, 1146a26-b1).) All of them involve an impairment in the belief's role in further reasoning. Someone who is asleep can be said to know, but is not in a position to do anything with her knowledge. The learner, a fairly different example, is another way of pointing to the same general area. If I have just learned the Pythagorean theorem, I can reproduce it easily, but I am to some extent "merely saying the words"; I do not have a firm grasp of the theorem and cannot yet draw out its implications. Something similar can be said about insanity and drunkenness. All of these involve belief in a fact, qualified by the tenuous character of the belief and by limitations on the capacity to draw out its consequences. We can recognize these limitations independently of any particular view of *akrasia*. They get us to recognize that, no matter what we think about *akrasia*, we already recognize a variety of beliefs with impairment in further reasoning from them. *Akrasia* can then be given a place in that same category, and seen as involving belief that is impaired in the same way. The person in Aristotle's central example thinks he should avoid sweets. His belief is reached by his own reasoning, and he would report it when asked, at least in many moments. But the belief does not fully function as beliefs ordinarily do in the ongoing life of a person. To that extent, they are impaired.[65]

[65] Unlike madness, drunkenness, and sleep, *akrasia* does not usually involve a general or across-the-board impairment. In that respect it is disanalogous. Gosling (1993) insists on this disanalogy. I do not emphasize it here, partly because madness, drunkenness, and sleep are not the only analogies Aristotle draws. When Gosling says that "There are three examples of conditions of 'knowing' to which that of the akratic is likened: sleep, drunkenness and either rage or

Impairment of this kind does not stop a belief from being a belief. These people will still tend to report the belief when asked, reason from it in many cases, and guide their actions by it to a large extent. Just as the learner already believes the Pythagorean theorem, the akratic believes she should not act as she does. Refraining from the action is not the only way this belief can be reflected in her behavior. As we saw in Chapter 3, she can report the belief to others, show signs of conviction in it, and give evidence for it. She can even visibly attempt to refrain. She can do all of this, even as she ultimately eats the dessert she believes she should not eat.

One of the main ideas in this chapter is that seeing how someone can act against her own better judgment is not especially difficult, and not what is most problematic about *akrasia*. In this I follow some of Aristotle's leading commentators. In her chapter on *akrasia* in *Ethics with Aristotle*, Broadie writes: "*His* solution to *his* problem I shall argue to be, for the most part, straightforward and obvious, almost anticlimactically so"(1991, 267). J.L.Ackrill wrote earlier: "There is nothing at all to be surprised at if a man acts against knowledge which he has but is not attending to"(1973, 31). I think Ackrill understates the difficulty, since, in some cases of *akrasia*, the akratic person does attend to the better judgment. In those cases the inner conflict is especially sharp. But in general, the presence of a prohibitive belief is compatible with akratic action, and by itself it does not threaten the explanation of the action by a different and conflicting evaluative belief. This is a central, underappreciated consequence of the Conflict Constraint.

Aristotle does not think of the akratic person as reaching the conclusion that she ought to do what she does. But I think we can extend his analogy to see the akratic as having a similarly impaired 'worse' belief. If I eat sweets despite the danger to my health, I may do it with a belief in the value of occasional indulgence, despite my considered view that it is not worth the health risks. My doing this intentionally is quite different from a case in which I absent-mindedly reach for a slice of cake, stopping myself with amusement and slight alarm after the first bite. But it is not done with a fully functioning evaluative belief, either. My belief that I should indulge my sweet tooth this time can be impaired in the same way as the other belief. It too is limited in its availability for further reasoning. In

---

madness"(100), he leaves out learning. In saying that "those who have just learned something do not yet know it," Aristotle is in the middle of discussing "different types of having" knowledge. Since he has in mind people who have already learned something, he presumably means that they do have the knowledge in one way, though they lack it in another. Learning is, of course, quite specific. When we have just learned the Pythagorean theorem, without the knowledge having fully "grown into us," the impairment is not across-the-board but limited to a particular set of attitudes, just as in *akrasia*.

Though I do not think the disanalogy is problematic, it is worth noting. The disanalogy may help explain why, unlike someone who is mad, drunk, or asleep, the akratic person is responsible for what she does. I do not consider the further issue of responsibility here, since my main concern is to explain how the Identity View allow for *akrasia* at all. But what I say does raise a more general question about responsibility for conflicting beliefs and intentions. Someone with severe *akrasia* in a wide range of circumstances may no longer be fully responsible for her attitudes; someone so full of conflict is in that respect like someone who is drunk or insane.

making further decisions, or answering questions about what I should do and why, I would often abandon my belief, and return to my more considered judgment. The impairment can be seen especially clearly if we imagine me "on the fence," with health losing my inner struggle. Here there is in a way little difference between *akrasia* and *enkrateia*, or strength of will, in this particular case. A fleeting thought can make the difference in what someone does in such a situation. Whichever option I take, both of my beliefs are to some extent unstable and less than fully functioning. I cannot fully accept the implications of either belief. Conflicting normative beliefs are always both impaired in this way; they, so to speak, impair each other. This does not show that I lack either of them. It is simply what having conflicting beliefs is like.

So far, I have begun to say how we can see conflicting normative beliefs in cases of akratic action and intention. But there are still at least several distinct and important concerns that can be raised about the Conflicting Belief View. In the next section, I continue to develop the view by responding to some natural objections.

III. Objections and Replies

1. Is the View Explanatory?

One might think that pointing to conflicting beliefs could not in principle provide an account of *akrasia*. It would not explain *how* such conflict is possible; so how would it explain how action or intention can be akratic?

I think this objection is right in wanting further explanation, but wrong to want it here. The Conflicting Belief View is not meant as a conception of akratic action or intention. It is a particular thought about akratic intention, which follows from the Identity View and can seem to be a problematic consequence of it. The Conflicting Belief View treats akratic intention as one species of a broader genus: conflict between beliefs about what one ought to do. Its explanatory ambitions are limited. My goal is not to explain everything one wants to know about akratic intention, but to defend a necessary condition on it. The challenge is to explain how, given that we intend akratically, the Identity View could be true.

On the other hand, if action on a normative belief is not in itself puzzling, the Conflicting Belief View suggests a way of reducing one problem—how can action or intention be akratic?—to another—how can we have conflicting normative beliefs? We are then left with one problem where there were two, and see why an understanding of akratic intention depends on an understanding of normative conflict between beliefs. That, I think, is genuinely explanatory.

Most importantly, the explanation is enough to address the problem *akrasia* poses for the Identity View. The problem was that the view seems to rule out the possibility of *akrasia*. The answer is that it does not. It can explain *akrasia* by a kind of conflict in belief. Most of us already agree that such conflict is possible, though we may not yet fully

understand how. This, I am arguing, is the beauty of noticing the possibility of such conflict, or accepting the Conflict Constraint. The constraint is minimal enough to accept, powerful enough to dispel doubts about *akrasia*, and interesting enough to get us thinking about how it can be true.

2. Can Agency be so Conflicted?

We often act akratically. If we were riddled with conflicting beliefs every time, would we be rational beings at all?

The problem is not just about the number or frequency of akratic actions and intentions, though it is tempting to put it that way. A mere appeal to numbers would be easier to answer. We act, intend, and believe against our own better judgment hundreds or even thousands of times; but how many times are we in accord with it? Our internally unchallenged beliefs and intentions may number in the millions or billions, if they can be counted at all. The ratio of conflicted to consistent cases may be impossible to determine or even estimate with any confidence. Nor is it clear what ratio would raise a problem.

The deeper difficulty is in the picture of a person as essentially disunified. Someone with conflicting beliefs can be hard to identify with, and to some extent hard to see as a person. Someone so unable to make up his mind begins to seem like he does not have a single mind at all.

In a later essay, "Paradoxes of Irrationality," Davidson considers this problem. On his view, as he puts it: "If we are going to explain irrationality at all, it seems we must assume that the mind can be partitioned into quasi-independent structures"(1982, 300). For him, these "parts of the mind are in important respects like people, not only in having (or consisting of) beliefs, wants and other psychological traits, but in that these factors can combine, as in intentional action, to cause further events in the mind or outside it"(290). They are, in other words, "organized elements, within each of which there is a fair degree of consistency, and where one element can operate on another in the modality of non-rational causality"(301). If a person is someone who moves rationally from one attitude to another, bringing various beliefs and intentions to bear on each other, then someone with flatly conflicting attitudes begins to look like multiple people, each with her own view and agenda.

Though much of Davidson's essay depends on details of his own theory of *akrasia*, his view about partitioning the mind does not. He defends it as a consequence of *any* view that takes the possibility of *akrasia* seriously. The subdivisions he has in mind are, strictly speaking, "not…independent agents" but "constellation[s] of beliefs, purposes, and affects"(303-4), each of which is internally consistent and can give rise to action, but without being able to stand in rational relations to the others. As Davidson puts it, "The breakdown of reasons-relations defines the boundary of a subdivision"(304). A mind is subdivided in this sense every time someone acts akratically. In every case of *akrasia*, and on every conception of it, the reasons-relations between a 'better' judgment and the person's action have broken down. The akratic is like multiple people to that extent, but

only to that extent. She is a person with constellations of attitudes that resist and undermine each other.

A concern about widespread conflict is partly a concern about the possibility of *any* conflict. We lose our picture of fully unified agency as soon as we accept the Conflict Constraint. Beyond this point, increasing disunity can also be increasingly disconcerting. As Davidson puts it, "It is a matter of degree. We have no trouble understanding small perturbations against a background with which we are largely in sympathy, but large deviations from reality or consistency begin to undermine our ability to describe and explain what is going on in mental terms."(303). The deviations may look somewhat larger when we think of *akrasia* as involving conflict in belief. But the difference is not in allowing that we do have conflicts in belief. Any view that satisfies the Conflict Constraint does that. Nor is the difference in the number of akratic cases, which every view already sees as involving a kind of disunity. The difference is in seeing a 'partition' within belief more often than other views do. This does sharpen the disunity, but there is so far no reason to think that it makes the difference between someone who is recognizably an agent and something that is not.[66]

---

[66] It might still lead us to shift the border line. Someone so irrational as to be almost insane may fall more easily into insanity. The scholastic view is committed to an added contradiction, and as Davidson puts it, "inconsistency breeds unintelligibility"(1982, 303). It would be helpful to look at such cases in detail; I am inclined to think that the added contradiction still would not make much difference in the extent to which we can see someone as a person, and that the difference, if any, would be in the right direction, theoretically speaking. Seeing all the contradiction would help us appreciate the full extent of the irrationality.

Explaining *akrasia* by conflicting beliefs does not move us from a properly unified picture of life as a person to an unrecognizably disintegrated one, but I think it does help raise the right questions about the unity of agency. These begin to emerge in some of Davidson's less guarded remarks about mental partitioning. He says at one point: "To constitute a structure of the required sort, a part of the mind must show a larger degree of consistency or rationality than is attributed to the whole"(300). Greater consistency in the parts is not obviously a consequence of the breakdown of rational relations between them, but it does clarify the puzzle. An akratic can reason from both her 'better' judgment and her 'worse' one; she may think she had better not cheat, or smoke, or steal, and still concoct elaborate schemes for getting away with her next violation. People are in conflict, and some parts, projects, or constellations of attitudes may be more consistent than the whole. Pointing to the conflicts does not unrecognizably undermine the unity of agency, but it may help us see how to study it.

Davidson's largest hint for further study comes in a footnote: "I have nothing to say about the number or nature of divisions of the mind, their permanence or aetiology. I am solely concerned to defend the idea of mental compartmentalization, and to argue that it is necessary if we are to explain a common form of irrationality. I should perhaps emphasize that phrases like 'partition of the mind', 'part of the mind', 'segment' etc. are misleading if they suggest that what belongs to one division of the mind cannot belong to another. The picture I want is of overlapping territories"(1982, 300n6). Though I leave out this complication in the text, tying *akrasia* to conflicting beliefs more urgently raises the question of the nature of the 'divisions', and how they can 'overlap'.

Instead, it brings out an advantage of explaining *akrasia* by conflicting beliefs. If we accept the possibility of conflicting beliefs in general, and start to see them here, we have helped to explain why *akrasia* is so puzzling. It is puzzling, at least in part in the same way that conflicting beliefs are puzzling. It is puzzling that one can be so disunified, and still be a single person.

3. Always Conflict?

Even if so much conflict is possible, is it really plausible to think that *all* akratic intention involves conflicting normative beliefs?

When one doubts that *akrasia* must involve conflicting normative beliefs, the alternative must be that we can be akratic without having both of those conflicting beliefs. At least one of the two conflicting beliefs must then be lacking. The doubt does not imagine the 'better' or prohibitive belief to be lacking. That belief is essential to the phenomenon of acting and intending *against* one's better judgment, or as we believe we ought not act. So the doubt must be about the 'worse' belief. One doubts whether the person must believe she ought to act as she intends to.

This doubt is not specific to *akrasia*. To be specific to *akrasia*, its guiding idea would have to be that an action may not be motivated or accompanied by a normative belief, *when the person has a conflicting belief*. Why would this conflicting belief make the difference? The thought may be that one cannot believe one ought to perform an action that one believes one ought not perform. But to think this is to violate the Conflict Constraint. It is to rule out the possibility of conflicting normative beliefs in cases of akratic intention.

If the objection does not depend on the presence of a 'better' belief, it is not an objection to the guise-of-the-good view's account of *akrasia*. It is a general doubt about the view. Rather than threatening the account of akratic intention, or rendering *akrasia* distinctively problematic, it expresses skepticism about the view as a whole. In that case, the way to address the objection is to take on the larger issue of which this chapter treats one smaller part: to think through the motivations and difficulties of the view, and see whether there is a substantive and defensible version of it. There is no distinctive problem about *akrasia* here.

4. The Error Attribution Problem

According to Aristotle (1984: I, 689), what "originates movement" is "either the real or the apparent good." But according to *us*, what we ourselves do is often not good, and not what we ought to do. *Akrasia* seems puzzling largely because, when we act akratically, it seems clear to each of us that we do and intend what we do *not* believe we ought to do. It then seems that the Identity View flies in the face of our own experience.

The view seems forced to attribute widespread error to people about their own beliefs. I call this the Error Attribution Problem. [67]

---

[67] Raz (2010) discusses this problem. I think his response is not very successful, but it helps to illustrate the one I go on to make.

Raz starts with two "apparent counterexamples"(2010, 112-3):

> A. *The miner*: The management proposes to close the colliery. The miners vote on whether to accept the proposal and the redundancy pay that goes with it or to oppose it. You talk to one of the miners: 'You are voting to stay put.' 'Sure,' he says. 'So you must have some hope [of keeping the mine open].] 'No hope. Just principles.'
> B. *The fish*: Sitting in the bath, Johnny…says, 'I am a fish' and beats the water with his open palm (presumably pretending to flap it with his tail). 'Why did you do that, Johnny?' 'That's what fish do.'

Assuming that they involve intention, these are apparent counterexamples to the Identity View, because "It may be difficult to get the miner or Johnny to acknowledge that there was value in the action. The miner may insist that his vote does no good….Johnny…was just playing…" (113). It seems the Identity View "must…attribute to the miner and Johnny…mistakes about their own beliefs"(114).

Raz makes two responses. "First, the notion of 'the good' or 'value'…is not to be confused with the concepts that are normally expressed by ordinary use of these terms"(114). He adds that there is "no point" trying to describe the "broader" technical notion, since "it is familiar from the writings on the subject," and in any case there is an "absence of agreement about its nature"(114).

Second, Raz says, his guise-of-the-good view "does not assume that agents capable of intentional action must have the concepts used in stating the [Guise-of-the-Good] Thesis…, nor…that they believe that these concepts apply to each of their intentional actions. It assumes [only] that they have a belief that…can be truly characterized as a belief that the action has a good-making property"(114).

I think both of these responses are better avoided. If a guise-of-the-good view did not use an ordinary concept, it would be hard to see what the view was claiming, how ordinary examples could be evidence for or against it, or how the ordinary concepts are related to the technical one. With neither a non-philosophical counterpart nor "agreement about its nature," the central notion would hardly be familiar enough. Most importantly, it is hard to see how Raz can think people would not need to have or use the concept used in stating the view, since in his second response and elsewhere, he uses the concepts in describing the content of the person's beliefs. One loses one's grip on what he thinks the view attributes to the person.

The miner and Johnny, it seems to me, both see what they do as good in a broad but still ordinary sense. The miner is acting on principle; if asked "What's the good in voting?", he may well say: "Principles." We see play as good; if Johnny does not, nothing brings that out. The miner and Johnny are very weak counterexamples; it is somewhat puzzling why they even *seem* to Raz to be counterexamples. I suspect that Raz leans toward identifying being good with producing some good. The miner and Johnny are both being in a way unproductive. The miner may insist that his vote "does no good," and both see no "value" in their actions (113). Raz's language is heavy with talk of consequences; without that, it is easy to think of principles for the miner, and play or fishhood for Johnny, as what these people see as making what they do better than the alternative. My focus on 'should' or 'ought' rather than 'good' helps to avoid this slide to focusing on good consequences.

It is telling how difficult it is to think of a good example to motivate the Error Attribution Problem. Raz's miner and Johnny are good examples for asking, as Raz goes on to, what the

Does pointing to a conflicting belief avoid the problem? People seem to often experience themselves doing and intending what they clearly *do not believe* they ought to do. People even say so. They say: "I don't think I should be doing this." The Identity View says that when they intend to do it, they *do* think they should be doing it. Why should we believe the Identity View, and not the person?

The Identity View attributes fewer mistakes than it may seem to. Our grammar makes the problem seem worse than it is. "I don't think he's being a good friend" does not usually report an *absence* of belief. It reports a belief. Normally, the speaker *does* believe that the "he" in the statement is *not* being a good friend. "I don't think we should invite him" is not an expression of indifference or hesitation, but a way of saying "Let's not invite him"; "I don't think I should be doing this" usually means "I think I shouldn't be doing this." As before, this is consistent with the Identity View. The view only holds that the person must *also* have a contrary belief.

We can imagine someone saying, "Trust me, I don't believe at all that I should be doing this." It is a point in favor of the Identity View that such assertions are rare. As Aristotle (1999, 1146b36) put it, such evaluative clarity without a practical change of mind "seems extraordinary." Even when someone does say this, it can often be insincere. It can also be a grammatically misleading expression of strong disapproval, rather than lack of approval. In other cases, the person might be right, because the 'action' is not intentional; it might be a bare reflex, or the involuntary cursing of someone with Tourette's syndrome, or a more complex compulsive movement, as in alien hand syndrome, of which the person is an alienated observer. Saying, "Trust me, I don't believe at all that I should be doing this" *can* be a lie, or a joke, or an expression of self-blame, or it can be straightforwardly true and unproblematic because the action is not intentional.

I think we should accept that it might not be any of these. The Identity View is threatened if it is forced to see people as systematically mistaken about their own beliefs. It does not have to say that we are *never* mistaken in this way. Such infallibility is unlikely, and a view that denies it is not at a disadvantage. So I think we can and should say that, in some special cases, we can be mistaken about what we believe.

Here, acknowledging the possibility of conflicting beliefs helps in two ways. First, it helps us see that the Identity View attributes error not in attributing a belief to oneself, but only in *denying* that we have a belief. If we seem to ourselves to obviously have some normative belief, the view will never disagree with us. This makes the range of attributed error smaller, by limiting the attributed error to error in denial.

Second, such mistakes are especially likely in the presence of conflicting beliefs, either of which can make the other easier to miss or deny. Some of us may be assuming when we act or intend akratically that, since we believe that we ought not do something, we do not believe that we ought. Even those of us who think conflicting beliefs are

---

argument *for* a guise-of-the-good view might be. But they are not clear examples of people to whom the view attributes a mistake. In the text I in effect argue that, to construct a clear example of error attribution in intention, one must build in so much that error does seem likely.

possible tend to ignore the possibility in our own activity. Pointing to the possibility of conflict offers an explanation of our mistake.

In sum, the Error Attribution Problem is powerful when the attributed error is extremely widespread. The Identity View attributes error in a small range of special cases. That, I think, is the right view. We are neither systematically mistaken, nor always aware of what we ourselves believe.

5. The Asymmetry Problem

We often describe *akrasia* as action against one's "better" judgment. But as I have described it, it seems to include action against *any* judgment—or, in the case of intention, *any* intention in a case of conflicting beliefs about what we ought to do. If we believe we ought to go to the beach, and believe we ought not go, anything we do will conflict with something we believe. To put it yet another way: neither belief seems in any way singled out as the "better" one. Both are treated as on the same footing. Explaining *akrasia* in terms of conflicting beliefs can then seem to leave out an essential *asymmetry* in an akratic's relations to her 'better' belief and her conflicting belief that she ought to act as she does. I call this the Asymmetry Problem.

By itself, pointing to conflicting beliefs says little about a symmetry or asymmetry. It does rule out one obvious possible asymmetry: that one potential action is believed to be what we ought to do, and another is not. But otherwise it leaves the issue open to the other features we might have in mind when we talk about a 'better' judgment. We often speak of the 'better' judgment as a more considered or "all things considered" judgment, which takes into account the full range of relevant considerations.[68] The 'better' judgment, in this sense, takes a wider view. As Davidson put it: "A judgment that, all things considered, one ought to act in a certain way presupposes that the competing factors have been brought within the same division of the mind"(1982, 301). An all-things-considered judgment brings one's views of all the competing factors into rational relations with each other, while an akratic one will tend to consider some and ignore others.

Because the emphasis on conflict is not yet a commitment to any particular view of asymmetry in *akrasia*, it can also allow that the 'better' judgment is better in other ways. The notion of a 'better' judgment is taken from its ordinary usage, which might not always be unequivocal. We can sometimes have in mind that a 'better' judgment is the one that we, on reflection, more strongly endorse or identify with. At other times, we might mean a considered judgment that is 'considered' in a somewhat different sense: considered more thoroughly, not only by taking a wider range of considerations into account, but by considering them more attentively or reflectively. We can also sometimes mean the judgment that is more reasonable, or more likely to be correct, and in that sense actually better. Though the range of evidence taken into consideration is *one* central dimension of

[68] That Davidson, for example, identifies the "better" or "best" judgment with judgment "all things considered" or "everything considered" is already clear in the opening sentence of his (1980a). See also the rest of that essay, and (1982, 294-7).

asymmetry, it is worth remembering that the scholastic view can allow for any of these other asymmetries.

In practice, the degree of 'asymmetry' can vary. The borderline case would be near total symmetry. In such a case, there is little difference between *akrasia* and its complementary state, *enkrateia* or strength of will. The akratic action is then not very strongly akratic, and the asymmetry is not crucial. Choosing a place to live, or choosing to have an abortion, may often happen this way. At the other extreme is a fully considered judgment giving way to the sudden appeal of a 'lesser' reason. Reading about Pluto on Wikipedia on a deadline evening can take this form. But to see an asymmetry, we need to think of the 'better' judgment only as the *more* or *better* considered one. It does not have to be *all* things considered. In practice, even our considered judgments do not consider everything that is relevant.

To illustrate these differenes, consider a more complex example of *akrasia*:

> *Family Values*. James is opposed to same-sex relationships. His considered belief is that we ought to "do what is good for our species, and not go against nature." He is also vegetarian, but has not thought about why, and is taken aback when the connection to his other views is pointed out to him. Because he believes that eating meat is natural and good for our species, he decides to give up vegetarianism. But that evening, he cannot bring himself to try meat.

James has decided that he ought to eat meat, but he akratically refrains. His case is complicated. His stated belief is itself not quite consistent; depending on how he understands it, letting nature run its course can conflict with doing what is good for *our* species. Nor is it an all-things-considered belief. James has not considered everything relevant; he had not yet thought about eating animals in this context, and there may be other things he has not thought about. His stated belief might itself be akratic. He might have a deeper belief in the sanctity of life, and in the importance of letting each living creature live and flourish without interference, which comes out vividly in many other contexts. In that case his stated principle is more local than he thinks, and may be little more than a rationalization of his homophobic feelings. We do not know exactly how much he has considered or where his deeper commitments lie. What we know is that he has two conflicting beliefs, and that his relation to them is probably asymmetrical.

If we believe the asymmetry is essential to *akrasia*, we can slightly revise the definition of *akrasia* to incorporate it. To *act* akratically is then to intentionally do one thing, while holding a more considered belief that one should not do it. To *intend* akratically is to intend to do one thing, while holding a more considered belief that one should not do it. On the Identity View, a conflicting belief can explain the intention and the action, while the "more considered" status of the other belief explains the asymmetry that is distinctive of *akrasia*. With or without this addition, the Identity View can allow for a wide range of *akrasia*.

IV. The 'Better' Intention

On the Identity View, intention is normative belief. My intention to eat dessert is a belief that I ought to. By the same token, my belief that I ought not eat dessert is an intention not to eat it. In intending akratically, I have both of these conflicting beliefs. By the same token, I have both intentions. I believe I ought to eat dessert, and believe I ought not to. I intend to eat dessert, and intend not to. So in cases of akratic intention, the Identity View is committed not only to attributing conflicting beliefs, but to attributing conflicting intentions. In this section I begin to make this attribution plausible.

Akrasia tends to involve an *inner conflict*. This is not true of every kind of irrationality or wrongdoing. Egoism and reckless disregard for one's own future are often quite wholehearted. But failing to do what one believes one should do typically involves a struggle with oneself. The contrasting state, *enkrateia*,[69] is often thought of as resoluteness or determination, and someone lacking these qualities might be described as wavering, hesitant, or torn. To gain a better understanding of *akrasia*, it helps to look at what kind of inner conflict is involved, what it is like to go through it, and how such cases are related to similar ones in which the inner conflict is different or lacking.

You believe that you should jump off the high diving board, and you know you have to climb up the ladder to do it. But when your turn comes, your fear of heights wells up and you 'chicken out'. Instead of climbing and jumping, you step away. But then you get mad at yourself, and get right back on the end of the line. You might even repeat the cycle over and over again.[70] This is a recognizable case of *akrasia*: you believe you should climb the ladder and not walk away, but instead you walk away. And looking at the example, one can see intuitively the place for conflicting intentions in explaining *akrasia*. What makes you fail to climb? You are torn between your initial decision and your overpowering fear. What makes you akratic, it is natural to think, is that you are intending both to climb and not to climb, both to jump and not to jump, even though you know you cannot do both. The heart of your *akrasia* is that you are being contradictory.

To see this, it helps to look closely at two examples in Christine Korsgaard's essay "The Normativity of Instrumental Reason." In the first (1997, 227),

---

[69] I do not consider *enkrateia*—continence, self-control, or "strength of will"—in this chapter. As far as what I say goes, it can be thought of either as acting on a better judgment despite a conflicting intention, or as acting on a better judgment despite a conflicting desire. I say nothing to decide this here, and nothing about desire or struggles against desire. This is unorthodox, but it is part of the explanatory strategy, and I think an advantage of it. If conflict in intention and belief explains *akrasia*, then any view of intention formation and desire will be compatible with the explanation, though not required for it.

[70] This is a variation on Korsgaard's (1997, 228-229) roller-coaster example.

> Howard…must have a course of injections, now, if he is going to live past fifty. But Howard declines to have this treatment, because he has a horror of injections.

Korsgaard gives this example to make a point about other principles—principles not about doing what one believes one ought to. But it is important to see the different ways Howard could be related to *akrasia*. He declines because of his horror. *How* does the horror lead him to decline? One possibility is that, as Korsgaard suggests, "avoiding the injections is what he wants most"(227). Howard's refusal of the treatment may be wholehearted and stable—a decision in the normal sense. If he thinks the choice is obvious, he may never seriously consider getting the treatment. To him, living into old age is simply not important enough to be worth going through such an emotional ordeal. He would rather have peace of mind now than extra years of life later. So he thinks it is good to avoid the treatment, and he unwaveringly leads the life he has chosen: he calls his doctor to say no, explains the situation to his family, and plans for an early death. If this is what he does, he may be shortsighted, a bad father and husband, and even irrational. But he is not akratic.

In a second, more likely case, Howard intends to act on his considered judgment in favor of getting the injections. But his fear overpowers him. He just can't go through with it. One week he drives himself to the doctor's office, circles around the doctor's block a few times, parks the car, gets out, stands on the sidewalk for a few terrified minutes, and drives home, missing the appointment. The next week he gets himself into the waiting room through a huge effort, but then apologizes and runs out. The week after that he cannot even make it out of his house, because his fear stops him at his front door. He never does get the treatment.

There is also a less likely third possibility. Howard could be like Jeremy. In another example of Korsgaard's,

> Jeremy settles down at his desk one evening to study for an examination. Finding himself a little too restless to concentrate, he decides to take a walk in the fresh air. His walk takes him past a nearby bookstore, where the sight of an enticing title draws him in to look at a book. Before he finds it, however, he meets his friend Neil, who invites him to join some of the other kids at the bar next door for a beer. Jeremy decides he can afford to have just one, and goes with Neil to the bar. When he arrives there, however, he finds that the noise gives him a headache, and he decides to return home without having a beer.[71]

This example too can be seen in different ways. Jeremy might conceivably have made a normal, reflective choice every time, for good reasons. Or he might just be especially fickle, often acting on whims. But suppose it's even worse: as Korsgaard imagines him, he

[71] (1997, 247n64). I quote Korsgaard's first use of the Jeremy example, though she later repeats it (with slight variations) and gives it a more central place; see her (2008, 116-7) and (2009, 169). Though Korsgaard does not compare Howard with Jeremy, she does proceed by considering sets of three related but contrasting cases, so my discussion follows hers in method if not in substance. See her own discussions of Howard and Prudence in her (1997, 227-229 and 236-237), respectively.

is "almost completely *incapable of effective action*"(2009, 169). Every time something else catches his attention, Jeremy suddenly loses or forgets his intention to carry out his current plan; his original commitment completely evaporates. His problem is then bad enough to be crippling or pathological, a debilitating absent-mindedness of the will. He may have to keep repeating his ends out loud in order not to forget them.

It is worth bringing Korsgaard's examples together to consider Howard's relation to Jeremy. Jeremy keeps forming intentions but failing to pursue them. We can imagine a variant of Howard's case, in which he himself has a similar problem. Howard might see or imagine the needle and have his considered intention to live a long life suddenly disappear. It could be that, like Jeremy, he has this kind of problem every day. Or perhaps just this one fear of his is so strong that it makes him simply forget everything else. Either way, he just drops his intention altogether. There are now three possible Howards: one who intends only to avoid the needle, one who intends to live a long life but is overpowered by fear, and one who 'forgets' his intention to live a long life as suddenly as he starts to feel his enormous fear of needles. They are, in other words, a determined Howard, a torn Howard, and a jolted Howard. All of them refuse the injections, but only the second one has an inner struggle.

It might be more natural to picture Howard in the second of the three cases—especially compared to the third, and especially when thinking of *akrasia*. The third case, like the case of Jeremy, is not a case of *akrasia* at all. *Akrasia* is not a failure to intend what we believed we should do in the past. Though suddenly giving up an end for no reason might be irrational, this irrationality, like that of the first, determined Howard, would not be akratic. The torn Howard, whose intention of living a long life loses an inner struggle, does act akratically, and phobias in general provide many classic cases of *akrasia*. The crucial point here is that the vivid case of *akrasia* is precisely the one in which Howard *does* still have the intention that he fails to pursue. His case suggests a general picture of *akrasia*. When we act akratically, we are led astray by fear, laziness, shyness, depression, anger, lust, envy, pride, greed, or some other desire, emotion, or vice. But we still have the 'better' intention. We are like Howard, who intends to go to the doctor but also intends to stay home.

One might think that Howard's 'better' attitude does not count as an intention, since he does not manage to act on it. But he does manage to act on it. Though he does not do what he sets out to do—get the injections—he takes many of the means, such as clearing his schedule and driving to the doctor's office. Nor could his fully carrying out the intention be necessary for having it. That necessity is independently implausible; we are often cut off from carrying an intention out to the end. Intentions are not all executed. And if there has ever been a case of two conflicting intentions, it must be possible to have an intention which is not carried out. In such a case, it is impossible to complete acting on all of one's intentions. So I think that, despite Howard's failure, we can see two conflicting intentions at work in his akratic activity. He is one example of conflict in intention.

Determining whether *all* examples of akratic intention involve conflicting intention is more complex. The answer to this question depends in part on which characteristics

should be seen as essential to intention. If normative belief is one of them, for example, the answer is yes. If not, the answer may be no. Attributing conflicting intention across all akratic intention thus goes hand in hand with developing a general conception of intention. This is the project not of this section, but of the entire dissertation. For now, what I have argued is more limited. I have argued, first, that akratic action and intention offer no immediately compelling counterexamples to the Identity View; and second, that the Identity View can shed light on some of the details of *akrasia*. I will return to the general conception of intention that has begun to emerge here in Chapter 7. So far, we have seen some of the details of the Identity View's treatment of one particular problem case.

In rejecting guise-of-the-good views, David Velleman writes that he hopes "for a moral psychology that can make room for the whole motley crew" of acting creatures (2000, 99). Michael Stocker writes in his essay "Desiring the Bad": "Philosophical theories…have depicted the psyche, especially the interrelations between motivation and evaluation, as far too simple, far too unified, and far too rational"(1979, 739). I think that some of our theories have depicted evaluation itself as far too simple, far too unified, and far too rational. Even the theories that explicitly allow for the possibility of conflicts in evaluation rarely do justice to the details or the metaphysical implications of those conflicts. In this chapter, I've tried to make more room for the motley crew of acting creatures by insisting on the complexities of evaluation, and especially of normative belief. And I've argued that it is naïveté about the coherence of our normative beliefs, not about their connection to intentional action, that we should be exorcising from our understanding of ourselves. Once we do that, even an ambitious view like the Identity View can become defensible. I think the lesson is more general: that a necessary connection between intention and normative belief can allow for and unify *all* of the phenomena of intentional action. What I have given in this chapter is one piece of the argument. I've argued that an understanding of the conflicts in our normative beliefs can offer a general account of akratic action under the guise of the good.

Chapter 5: Motivation without Evaluation

In Chapter 4, I considered one central kind of counterexample to the guise of the good. On what I have been calling the Identity View, intention is a kind of normative belief. To intend to do something is to believe one should do it. But when acting akratically, we intentionally do something, and usually intend to do it, while believing we should *not* do it. Intention and normative belief seem to point us in different directions. We endorse an alternative action such as eating more vegetables—or at least a refraining, such as not eating dessert. We intend one thing and endorse another. And so, it seems, the intention cannot itself be our normative belief. An intention to eat dessert cannot be a belief that one should *not* eat it. To resolve this problem, I argued that the intention can be understood as a second normative belief that conflicts with the 'better' belief that makes the intention and action akratic.

Akrasia is not the only important counterexample. In one way, it is not even the most direct. It is natural to think we can intend to eat dessert, without having a normative belief about it at all; or believe we should get out of bed, but not intend to. It seems we can simply have either an intention or a normative belief without having the other of the two. There are many apparent cases of this kind, and, I think, none of them can be understood in terms of conflicting beliefs. They are cases not of apparent conflict but of apparent indifference: the normative indifference to *which* bowl of ice cream to eat for dessert, or the motivational indifference of the inability to get ourselves out of bed. These are the topic of this chapter and the next one, beginning here with intention without normative belief.

The clearest apparent examples of intention without normative belief are ones in which we seem entirely unable to form a normative belief. Faced with two similar bowls of ice cream, we can find ourselves unable to see either one as the one we should take—but then we take one nevertheless. In other cases, we just cannot see how to compare the options. Shall I care for my sick mother or fight in the resistance? These may not be equally worth doing, but if I cannot decide which one I should do, I may have to decide to do one without believing it is what I should do.

I will begin in §I with the simplest cases, in which two or more alternatives are believed to have no relevant differences between them. In §II, I consider four natural responses the Identity View could make to such cases. One can: (1) deny that there are any such cases; (2) concede that they cannot be successfully resolved; (3) insist that we simply select at random; or instead, (4) say that we simply let our attention fall on one of the alternatives, and take that one. I will argue that all of these responses fail to account for at least some intention in such cases. In §III, I describe a different view: that in a typical intentional resolution of such a case, we decide to act non-intentionally. We decide to simply 'pick', in a particular sense of that word: we decide to let our activity continue

without intention, until one of the available alternatives emerges as the one we ought to take. I will argue that this response succeeds whether the others fail. §IV extends the account to related cases of 'existential' choice and inability to compare alternatives. Throughout, the central idea will be that, when a person does form an intention without a belief favoring one alternative, that intention is still a normative belief. In many typical cases, it is an intention, and a belief that she ought, to act non-intentionally.

I. Buridan's Ass

The medieval Arabic philosopher-theologian Al-Ghazali (1963, 26-27) described a choice between two identical alternatives:

> Suppose two similar dates in front of a man who has a strong desire for them, but who is unable to take them both. Surely he will take one of them through a quality in him, the nature of which is to differentiate between two similar things. All the distinguishing qualities...like beauty or nearness or facility in taking, we can assume to be absent, but still the possibility of the taking remains.[72]

Faced with two pieces of fruit, Al-Ghazali's imagined man can find no difference between them on which to base his choice. They are equally beautiful, equally close and convenient, and, we can assume, equally tasty, healthy, and so on. But he is nevertheless able to take one. Al-Ghazali insists that we are able to choose between such alternatives.

Al-Ghazali's dates are an earlier variant of an example that became famous as Buridan's Ass. The imagined ass, or donkey, finds itself hungry midway between two equally sized bundles of hay. Unable to choose, it dies of starvation. The example, though not found in John Buridan's writings, is widely thought to have arisen as an objection to his view of the will as determined by reason.[73] If the will could only do what reason commands, the example suggests, it would be paralyzed by a situation in which reason could not command either option.

---

[72] The precise date of composition is unknown, though estimated by, e.g., Bouyges (1927, ix) at 1095. Al-Ghazali's lifetime was 1058-1111. In the text I use Van der Bergh's translation from Averroes (2008, vol. I, 18-23), a detailed twelfth-century commentary on and refutation of Al-Ghazali. For further discussion, see Rescher (1959), 146-150.

[73] See Rescher (1959, esp. 153-5) for discussion. Buridan lived from circa 1300 until shortly after 1358, roughly 250 years after Al-Ghazali. Al-Ghazali's example arose in the context of an attack on the analogous view to Buridan's but about divine will—the principle of sufficient reason for God's actions—and thus in defense of the rational inscrutability of God's actions. In each case, the example arises as a central counterexample to the guise of the good.

Although, as I go on to say, Al-Ghazali's is a more useful example, I do change it slightly. I ignore his talk of desire, adapting the example to focus on intention, and I will occasionally consider a variant in which the dates are not exactly identical.

The donkey dies; the man does not. But in each case, the lesson is the same. Neither donkeys nor people would actually die in such a situation. A theory that requires them to act as they believe they ought seems unable to account for this fact. Neither date, and neither hay bundle, is any better than the other; and, it seems, neither is believed to be better, or more worthy of choice, or what one ought to take, by the character in the respective story. If the donkey and the man could act only as they believe they ought, it seems, they would starve; they in fact would not starve; so they must be able to act other than as they believe they ought.

The same problem arises about intention. When we believe that we can intend only under the guise of the good, then, to borrow Michael Bratman's (1999, 220) words, "we have our Buridan problem. It seems that I can just decide on which bookstore to go to, while continuing to see each option as equally desirable." Bratman's earlier consideration of Buridan's Ass (1987, 11) already takes this view:

> I conjecture that we have an ability that is basic at the level of commonsense psychology: an ability to decide in the face of equidesirability….This has implications for coordination. If I want to coordinate with you, I need to know not just the desires and beliefs in light of which certain alternatives are seen by you as equally desirable; I also need to know your intentions.

Implicit in the idea of needing to *also* know someone's intentions is a denial that intentions can themselves be beliefs. The examples of the dates, hay bundles, or bookstores illustrate this idea. These are apparent cases of what Bratman helpfully calls "underdetermination by value judgment"(2007, 161).[74] Imagine the dates a short walk away; the man can form an intention to take the one on the left, walk over to it, and take it—apparently without believing that he ought to take that one and not the other one. If he could intend only by having a normative belief, it seems, the man could not intend to take either date; in fact he can; so he must be able to intend without forming a normative belief in favor of either date; and so an intention cannot in general be a normative belief. We can call this the Buridan Argument against the Identity View. I believe the argument fails, because its first premise is false. In this section and the next two, I will focus on such examples and the problem of how to account for them. I will try to show how the man can succeed, even if he can intend only by having a normative belief. If the Identity View can be reduced to absurdity, I will argue, it is not by examples like Buridan's Ass.

Al-Ghazali's example is not only earlier than Buridan's, but also clearer and more relevant. The introduction of a donkey raises questions about whether non-human animals can act intentionally, act at all, or have intentions, normative beliefs, or a will, which are not essential to this particular problem. And although death is a dramatic feature of the

[74] As the longer passage from Bratman illustrates, he believes not just that intentions are not reducible to beliefs, but that they are not, as many Humeans have thought, reducible to combinations of beliefs and desires. I leave out desires here, and focus on one immediate consequence of Bratman's view: that intentions are not a kind of belief.

example, and makes for a vivid caricature, it can obscure the issue. What needs to be shown in order to defend the Identity View is not just that the donkey will not die. What needs to be shown is that, if it forms an intention, it does so under the guise of the good. Al-Ghazali's example raises this problem directly. The man seems able to act intentionally, and form an intention in advance, without a corresponding normative belief.

Nevertheless, since later writers have mostly focused on Buridan, I will refer to these as "Buridan cases." It will be helpful to try to include a wide range of cases that raise the same basic problem. As I will use the term, a *Buridan case* is *a case in which a person has multiple available courses of action, which she believes to have no differences between them relevant for belief about which one she ought to take, and each of which she would choose over any available action outside the set*. A few terminological clarifications will help here. By "person" I mean a being capable of intention and belief, of any biological species or physical kind. "Multiple" allows cases with more than two potential actions, such as picking a card from a deck; there is no need to assume that Buridan cases must be binary. "Courses of action" leaves open the possibility of Buridan cases that are not cases of taking an object; they might, for example, be cases of sending garbage away for disposal to one of two equidistant dumps, or of singing one of a series of notes. The focus on "believed" difference is important; a mistaken factual belief about actually identical alternatives could easily allow a person to form a mistaken normative belief, while the relevant case is an apparent case of no normative belief at all. "Relevant" allows for obviously irrelevant differences: though cases of clearly identical alternatives are the clearest cases, a choice between dates arbitrarily labeled "8371" and "8713," or with beliefs about which one is farther west, would still present a problem for the guise of the good if the differences are believed irrelevant to belief about which one we ought to take. Such cases are naturally seen as lacking normative belief, but allowing for intention and intentional action. On the other hand, it does make a difference that the apparently identical options not be accompanied by another option that the chooser believes she should take instead of the identical ones. If Buridan's ass saw a bigger hay bundle closer than the others, or Al-Ghazali's man saw a third, tastier date, we would no longer have an apparently clear example of intention without normative belief. It would be easy for the man to form the belief that he ought to take the third date. This is why I demarcate the range of Buridan cases the way I do, though other kinds of case can be closely related.[75]

---

[75] We can give names to some closely related types of case.

A *binary Buridan case* is a Buridan case with only two options with no differences believed relevant for normative belief. Al-Ghazali's man and Buridan's ass are both binary Buridan cases.

An *objective Buridan case* is a case in which the alternatives in question are in fact identical in all relevant respects, not merely believed identical. Al-Ghazali's man and Buridan's Ass can be imagined as objective Buridan cases, or as cases of real but insignificant and imperceptible differences in size and shape. Objective Buridan cases are simpler, and the chooser's ignorance or mistake about an actual difference can be an unnecessary complication. But "objective Buridan cases" may or may not be Buridan cases. Al-Ghazali's man could be misled into believing that one of the dates is tastier than the other, even though they actually taste the same. He would

The man's taking of a date is an apparent case of *action without normative belief*. He takes, say, the date on the left, without thinking he ought to take the date on the left. If we think of his action as intentional, it is an apparent case of *intentional action without normative belief*. And if he forms an intention before acting on it, it is an apparent case of *intention without normative belief*. He might, of course, intend to take *a* date and believe he ought to. But the intention to take a particular date, like the intentional taking of that particular date, seems to have no normative belief corresponding to it. It seems that the man can intend to take the date on the left, and intentionally take it, without believing he ought to take that one. In this way the example applies to action in general, to intentional action, and to intention. For any of these, it can seem unclear how to defend a guise-of-the-good view when faced with such a case. The problem is especially stark for a view that identifies intention with normative belief. I will start with what I think are unsatisfactory responses to the cases, and use them to develop a defense of the Identity View.

II. Initial responses

1. Denying the phenomenon

Al-Ghazali (1963, 27) goes on to tell an imagined objector:

> You can choose between two answers: either you merely say that an equivalence in respect to his desire cannot be imagined — but this is a silly answer, for to assume it is indeed possible — or you say that if an equivalence is assumed, the man will remain for ever hungry and perplexed, looking at the dates without taking one of

---

then believe them to be different in a relevant respect, and could form a normative belief on the basis of that difference.

An *indiscernibility case* is a Buridan case in which the preferred options are not believed to be different in any way, rather than believed identical in every way relevant for choice. I will avoid talk of strict indiscernibility. When Al-Ghazali says that "all the distinguishing qualities... like beauty or nearness or facility in taking, we can assume to be absent," his wording can be heard as building indiscernibility into the case. But "distinguishing qualities," as he conceives of them, include extrinsic or relational qualities such as nearness to the man. They can then also include nearness to Mecca or to the north pole. If imagined in our world, the two dates will differ in geographical location in a way that puts one closer to some locations and the other closer to others. It is a substantive question whether genuine indiscernibility cases are possible, and whether one can have distinct potential objects of choice without any believed difference by which to distinguish them. But as before, this is not essential to Buridan cases, which can (and perhaps must) involve an apparently irrelevant difference between the options. Such irrelevant differences are possible, and the problem arises even when they are present. As Rescher (1959, 143) puts it, "Indiscernibility is not at issue here, but merely indistinguishability *qua* objects of choice, so that every known reason for desiring one alternative is equally a reason for desiring the others." Even without indiscernibility, deliberation toward normative belief can reach a tie, which seems breakable not by belief but only by action or intention. Buridan cases do not have to be binary, objective, or indiscernibility cases.

> them, and without a power to choose or to will, distinct from his desire. And this again is one of those absurdities which are recognized by the necessity of thought. Everyone, therefore, who studies, in the human and the divine, the real working of the act of choice, must necessarily admit a quality the nature of which is to differentiate between two similar things.

Both kinds of answer have been given, and we can consider them in order. First, one might wonder: can we really imagine such a case? Are any two options ever exactly identical, even from the point of view of a chooser? It can seem natural to simply deny the example. Montaigne seemed to do just that. He wrote (1877, vol. II, 381-2):

> 'Tis a pleasant imagination to fancy a mind exactly balanced betwixt two equal desires…. Nothing presents itself to us wherein there is not some difference, how little soever; and…either by the sight or touch, there is always some choice, that, though it be imperceptibly, tempts and attracts us.

Leibniz similarly wrote (1952, §49):

> Buridan's ass…is a fiction that cannot occur in the universe…. For the universe cannot be halved by a plane drawn through the middle…so that all is equal and alike on both sides…. There will therefore always be many things in the ass and outside the ass, although they be not apparent to us, which will determine him to go on one side rather than the other."[76]

Montaigne and Leibniz deny the phenomenon. And if there is no such phenomenon, there is nothing problematic for the guise of the good to explain.

Such a denial is on shaky ground, for several reasons. First, the first answer as Al-Ghazali put it was that the relevant kind of equivalence cannot be *imagined*; Al-Ghazali's reply was that it is possible to *assume* it. Neither Montaigne nor Leibniz deny this.[77] Even if there are never in fact any such cases, merely imagined examples, assumed to be possible for the sake of argument, would be examples in which we seem forced to think of

---

[76] Montaigne's view comes in his *Essays*, Book II, Essay 14, the one-page essay "That the mind hinders itself"; Leibniz's comes in a short passage in his *Theodicy*. Though this is not usually noticed, each combines the first two responses I mention; see below. The quotations from Montaigne and Leibniz also appear in part in Ullmann-Margalit and Morgenbesser (1977), 759-760, who take them to be examples of denying that there are situations of "strictly indifferent preferences"(759). But as both they and I go on to clarify, a situation can be one of "indifferent preferences" even when the options are not in themselves identical. See their discussion of Leibniz's "petites perceptions"(763).

[77] Leibniz adds: "Fundamentally the question deals with the impossible, unless it be that God bring the thing about expressly"(1952, §49). He considers it impossible, but not beyond God's power, and at any rate not inconceivable. Indeed, both Montaigne and Leibniz themselves imagine such cases and consider what would happen in them; see §III.2 below.

an action or intention as lacking any corresponding normative belief. Denying the phenomenon does not deny the possibility of imagining the phenomenon, and there is room to ask what a guise-of-the-good view can say even about imaginary cases.

Second, the differences Montaigne and Leibniz consider might help show only that the donkey and the man can avoid starvation. This would not show that their intentions are normative beliefs. As Montaigne and Leibniz put it, respectively, the differences between alternatives can be extremely "little," and "may not be apparent to us." Such imperceptible differences might influence intention and action, without working by way of normative belief. They might then leave Al-Ghazali's case just as it is. A slight difference can lead Al-Ghazali's imagined man to take the date on the left, without the man believing he ought to take that one. As with the donkey, what needs to be shown in order to defend the Identity View is not just that the man will not die. What needs to be shown is how, if he forms an intention, he can do so under the guise of the good. Imperceptible differences do not obviously show an action or intention to be under the guise of the good. They leave open the possibility that, if the donkey and the man do not starve, this is only because they can have intentions that are *not* normative beliefs.

Third, and relatedly, a simple denial that there are such cases tends to be unclear about what kind of example is a relevant counterexample. Neither the dates nor the bundles of hay are strictly, or exactly, qualitatively identical. Any two real dates or bundles differ at least imperceptibly in size and shape (though they can be imagined not to). The threat to the guise of the good is in the way such examples seem to be clear cases of action, intentional action, or intention without normative belief. It is enough if the two dates are identical as far as the chooser can see, with no perceptible differences between them. To create a problem, it is enough that the chooser can see no *relevant* difference. As we saw, Buridan cases can involve unnoticeable and even noticeable differences, as long as these present no basis for normative belief. Faced with two hay bundles labeled "8713" and "8731", the donkey would still be in trouble. The absence of a noticeable difference is just an especially clear example of alternatives without any difference that could give rise to a normative belief in favor of one of the alternatives.[78]

Fourth, and most importantly, the denial flies in the face of the apparent fact that there are Buridan cases. When we pick one dime from a pile, or card from a deck, or piece of candy from a tray, or box of cereal in a grocery store aisle, we seem able to pick

[78] There is an intermediate possibility here: a defender of the Identity View could insist that an imperceptible difference causes a normative belief, which is an intention. In other words, the difference could cause the belief itself, without being a reason for it. The normative belief might be held without reasons. I do not rule out this possibility here. But I avoid insisting that all apparently Buridan-like cases are resolved in this way. That is an unlikely strategy. For one thing, if someone does form a normative belief in this way, she could reflect on her belief, find no reason for it, give it up, and be back in the same situation. It is far from clear that imperceptible differences could always give rise to motivationally effective normative belief in such a brute way. Moreover, if they did, it would be hard to avoid the implication that the belief (and therefore the intention) is irrational, and that all of our responses to Buridan-like cases must be irrational. I consider the rationality of our responses to these cases in §III.4 below.

intentionally, and to form a corresponding intention in advance: "In a minute I'll take *that* card." And we seem to do this without believing we ought to take *that* card, or this piece of candy. Of course, there will be related cases in which there is a clear best choice, and the chooser sees this. And if the Identity View is right, there can never be cases of intending to take a particular card without believing that one ought to take that card. But there are many such cases that seem to cast doubt on the Identity View. As Ullmann-Margalit and Morgenbesser (1977, 761) put it: "Supermarket shelves supply us with paradigmatic examples…. Having eliminated from among the rows upon rows of Campbell tomato soup cans the less conveniently accessible ones as well as the conspicuously damaged ones, you are still facing at least two cans neither of which is discernibly superior to the other(s)."[79] Cases like these still need to be accounted for. A denial of their existence is a desperate and unsuccessful move with a dubious theoretical motivation. The question remains: how do we manage to form an intention to take one alternative, when we can see no way to come to a belief about which one we should take?

2. Biting the Bullet

Al-Ghazali's second imagined response was that we do not manage it. Choice in such cases is simply impossible; "the man will remain for ever hungry and perplexed." Aristotle suggested such a view in passing in his mention of "the man who, though exceedingly hungry and thirsty, and both equally, yet being equidistant from food and drink, is therefore bound to stay where he is".[80] Buridan himself can be imagined to have made such a response, and he would have had respectable followers. Montaigne, while denying the phenomenon, accepts that it would have a disastrous outcome. "Were we set betwixt the bottle and the ham, with an equal appetite to drink and eat, there would doubtless be no remedy but we must die of thirst and hunger"(1877, vol.II, 381). Like Montaigne, Leibniz made both of the responses Al-Ghazali imagined, adding: "It is true that, if the case were possible, one must say that the ass would starve himself to death" (1952, §49). Concluding Part II of his *Ethics*, Spinoza (2002, 276) wrote of a similar example:

> I readily grant that a man placed in such a state of equilibrium (namely, where he feels nothing else but hunger and thirst and perceives nothing but such-and-such food and drink at equal distances from him) would die of hunger and thirst. If they ask me whether such a man is not to be reckoned an ass rather than a man, I reply that I do not know, just as I do not know how one should reckon a man who hangs himself,…[or] babies, fools and madmen.

[79] For the contrasting view that genuine 'motivational ties' are rare, see Mele (1991) and (1992a, 67-78). I partly agree with Mele in §III.5 below. Mele's discussion considers Buridan cases in the context of a discussion of motivational strength, and thus addresses a somewhat different problem.
[80] *On the Heavens*, Book II, Chapter 13, line 295b24 in the standard Bekker pagination. See Aristotle (1984).

Spinoza admits the puzzle; he does not quite know what to make of the example. But he thinks it no less real than infancy or suicide or insanity. In his picture of the world, there can be Buridan cases, but we are simply unable to choose between the alternatives in those cases.

This second response is what we would now call "biting the bullet." While the first response denies the possibility of Buridan's Ass, this one takes the donkey's plight as paradigmatic of what happens in such cases. It denies, not the existence of Buridan cases, but the possibility of their resolution—that is, of performing one of the candidate actions. But like the first response, the second one depends on a denial that is hard to sustain. There are again several reasons to resist it.

First, there is the lack of a clear rationale *for* the response. There seems to be no reason to accept that the donkey or the man will starve, except as the consequence of a theory. But we can ask: is it a consequence of our theory? I will soon explain why I think it is not, if the theory is the Identity View.

Second, to be convincing, a response like this would need to explain why examples like Buridan's ass seem so absurd. Spinoza accepts a view which for Al-Ghazali is "one of those absurdities which are recognized by the necessity of thought." Why would Al-Ghazali think this? Why would the donkey's plight strike people as absurd enough to tell against a general theory that condemns him to it? Spinoza's analogies can make this question more pressing. Babies, fools, and madmen all lack full rationality, intelligence, or judgment. Why would an ordinary adult be reduced to the state of a paralyzed donkey when put into such a choice situation? Spinoza's puzzlement here makes his own view less easy to accept. It offers no explanation of the apparent absurdity.

Third, and most importantly, this response, like the first one, seems to fly in the face of what we already know. If it is obvious that there are Buridan cases, it seems equally obvious that there are resolutions of them. As Al-Ghazali would insist, we do in fact choose between two dates, or between food and drink, or between two equidistant bundles of hay, or between many cans on a supermarket shelf. This is part of the power of Buridan cases. They seem imaginable and even widespread; and they seem resolvable. It seems clear that deliberation can reach a tie, without paralyzing the deliberator. It then seems that we can form intentions, and act intentionally, even when there is no way to form a belief that we ought to take the option we intend to take.

3. Randomization

In the passage quoted earlier, Al-Ghazali told his imagined objector: "You can choose between two answers." He was too quick to think there are only two. The Identity View does not have to deny either that there are Buridan cases, or that we can successfully resolve them.

Rescher (1959) offers an alternative possibility. According to Rescher, when faced with what he calls "the problem of choice in the absence of preference,…*random selection*

is the only reasonable procedure"(170). There is no defensible basis for favoring any particular alternative; so the remaining option is to select at random.

This too is not an easy fix. To begin with, what does "random" mean? If it means "arbitrary," which in turn means "without a reason," it seems that, by hypothesis, *any* solution will count as random. The response is empty unless it uses "random" in a narrower, more interesting sense. [81]

It is natural to think of random selection as selection using a randomizing device, such as a coin or die.[82] In a reply to Michael Bratman, Davidson (1985b, 200) points to a coin toss as a preferred method of resolution:

> [Bratman] correctly points out that sometimes we have to decide even when there are no obvious grounds for decision. But if there is reason to reach some decision, we find extrinsic grounds. Perhaps I flip a coin to decide. My need to choose has caused me to prefer the alternative indicated by the toss; a trivial ground for preference, but a good enough one in the absence of others.

A coin toss, Davidson thinks, can indicate one alternative to us, breaking the apparent tie.

But the use of random devices faces a regress of Buridan cases. *Which* random device shall I use? If I have a coin and a die, or five coins, I might see no relevant difference between them. How shall I determine which one to use? Do I use one of the devices I have for that too? Which one? The choice among random selection procedures can itself be a Buridan case, which we can be unable to solve except by resolving the same case in some other way.

There is an even worse regress, which arises even with one coin. Flipping a coin and watching it fall does nothing if the available actions are not matched to the possible outcomes of the coin toss. To use the coin, a person must first decide: should the date or hay bundle on the left be taken if the coin comes up heads, or if it comes up tails? But of course, *this* makes no difference. There is no reason for assigning left to heads that is not also a reason for assigning it to tails. We can call this the *matching problem*. Given a coin, or a die, or any other such randomizing device, Al-Ghazali's man would have to match its

---

[81] A more systematic development of this response could consider various prevalent senses of "random" and say which, if any, could offer a response to Buridan cases. I do not attempt this here. In what I go on to say, two related features stand out: (1) selection not for a reason; and (2) 'delegating' one's choice to some process other than deliberation.

[82] Rescher gives the credit to the 17th century British scholar and theologian Thomas Gataker for being "the first to suggest the employment of random-selection devices as a means of resolving the problem of indifferent choices"(1959, 156), and takes inspiration from him. It is worth noting that these devices do not need to be truly random in any statistical sense. We would not be upset, for example, to find out afterwards that the dice were loaded or the coin 'biased' toward heads. We just need to delegate choice to something that does not have to 'choose' for a reason. For the same reason, the use of a mental 'randomizer', discussed below, is not jeopardized by the possibility of implicit bias. From the deliberator's point of view, the alternatives are equally good, and it does not matter whether any randomizing process is, unbeknownst to her, weighted in advance toward a particular outcome.

outcomes to the alternatives it is meant to help him choose between. Rolling a die would not help; the man would still have to decide which outcomes of the die call for which alternative matching. Nor would it help to have a preference for, say, tails in coin tosses. Even if the man likes tails more than heads, which date should he match to it? If he is in a Buridan case, he will see no reason to match his favored tails to the left date that does not apply to the right date.

The matching problem is an especially vicious regress. It points to a Buridan case within the resolution of any Buridan case using a typical randomizing device such as a coin or die. It does not depend on the accidental presence of a second, equally good randomizing device. And it is not just a problem for the character in the story. It is a theoretical problem that threatens any appeal to randomization. As we saw, we do seem able to match act options to outcomes of a random process. We do seem able to flip a coin to choose which date to take (though, as the matching problem brings out, it is not so clear why we bother with the coin). What we cannot do so easily is say, in general terms, how we can resolve these cases without an intention that is not itself a normative belief. Since randomizing devices can create a matching problem, they do not provide an obvious general answer. To solve this problem, an appeal to randomization would have to say how we can use a randomizing device to resolve a Buridan case without already resolving a Buridan case in some other way.[83]

Rescher's solution is to adopt a random "*policy* of choice"(1959, 168). We can, for example, always take the date on the left, or the first-mentioned option in a list. The policy dictates what to do, provides a justification for one's choice, and, according to Rescher, avoids the regress. "When I make a choice among symmetrically characterized alternatives, I *can* defend it, reasonably, by saying, 'I chose the first mentioned (or the like) alternative, because I *always* choose the first-mentioned (etc.) in these cases…. This alone averts an infinite regress of random selections"(1959, 168).

Rescher's solution can be seen as a way of extricating oneself from a Buridan case through the formation of a normative belief. Al-Ghazali's man can adopt a policy of always taking the date on the left when faced with a pair of dates. Offered a choice of dates, he might then think: "I have a policy for cases like this. I should follow the policy. So I should take the date on the left." This normative belief can be seen as constituting an intention to take the date on the left. Appeal to a random policy is thus one way to defend the Identity View, while also explaining how our resolution of Buridan cases can be rational.

---

[83] The matching problem also raises the interesting question of why we bother with the coin. I leave aside here the utility of coins and dice, and of other randomizing devices, such as lottery machines, that might not create a matching problem.

For earlier mentions of regress problems for randomization, including what I call the matching problem, see Leibniz (1948, 488), quoted in Strickland (2006, 151); Ullmann-Margalit and Morgenbesser (1977, 769-70), and Stöltzner (2000, 28), a discussion of Neurath (1983). Mintoff (2001, 213) also points out the problem, rejects "various suggestions"(212) for resolving it, and concludes that it is decisive against views like the Identity View, but without considering the solution I go on to describe.

But this response too is unlikely to succeed. At best, it reduces the number of Buridan cases a person has to encounter. Al-Ghazali's man does not face a genuine Buridan case if one of the dates has a distinguishing quality—being on the left—which, given the man's policy, favors taking that one. But this only holds if the policy is already in place. If he does not already have a policy, he must adopt one policy rather than another. As Rescher puts it, "I *always* choose the first-mentioned (etc.) in these cases"; but the "etc." is crucial. I could instead choose the last-mentioned, or the one on the right, every time. It often makes no difference whether I adopt a policy of taking the one on the left, or right, or the first- or last-mentioned in a list. The adoption of a policy can then itself be a Buridan case. And when Al-Ghazali's imagined man, who has no such policy, must make his choice, the possibility of adopting a policy does not change the fact that, whether or not he adopts a new policy now, he must take one of multiple alternatives each of which he believes to have no distinguishing quality relevant to a normative belief in favor of one or another.[84]

The possibility of forming a policy thus faces the same general regress problems about any method of 'random' or 'arbitrary' choice. *Which* way of choosing shall I take as my policy? A policy of taking the left one? Or the right one? Or shall I flip a coin? A coin with heads matched to the left one? Or to the right one? A die with 1-3 for left and 4-6 for right? Or vice versa, or odd and even numbers? When I have both, shall I use a coin or a die? There are many ways to delegate one's choice of policy to a randomizing device, and the selection of one is itself a Buridan case. And if we use a randomizing device to select a policy, we again face the problem of matching outcomes of the device to the possible policies. The regress has not yet been stopped. Even when a policy can help, we have to select one. And without a solution to the matching problem, no coin can help.

If Rescher's solution can work, it must be supplemented by some way of selecting a random policy. A coin or die would, of course, face the same regress. But on his view, "This randomizing instrument may, however, be the human mind, since men are capable of making arbitrary selections, with respect to which they can be adequately certain…that the choice was made haphazardly, and without any 'reasons' whatsoever"(1959, 169).[85] A random policy can be adopted, Rescher might say, because of the human capacity for

[84] My criticism of Rescher's solution draws on Ullmann-Margalit and Morgenbesser (1977). They go on to point out that any policy will have to have contingency plans: we might take the first alternative in a list given orally, the leftmost spatially, the uppermost vertically, the first alphabetically in other situations, and so on, and higher-order policies may have to adjudicate conflicts that arise between these, becoming fairly complex and creating further Buridan cases at the meta-level. As they also point out, resolution of a Buridan case does not seem to require the consistency involved in *always* following a policy. Rescher's solution thus seems unfaithful to the ways in which we do form and act on intentions in Buridan cases. I leave out these other criticisms in the text, to focus on the regress problem.

[85] Rescher says this in a somewhat different context, while considering a policy of using a further, randomizing device in some cases. In other words, he is considering the content of the policy, not the way it is adopted. I adapt the point on his behalf to address the problem of the policy's adoption.

arbitrary mental selection. The human mind, or some capacity of it, can itself function as a randomizing device.

The appeal to a capacity for "making arbitrary selections" then seems essential for stopping the regress created by an 'external' device. But it faces two problems of its own. First, how do we make these selections? The process is so far completely obscure. Second, if we have this capacity, why do we need a policy? Instead of arbitrarily selecting a policy and following it every time, why not just arbitrarily select one of the two dates? The policy begins to look like an unnecessary middleman between the set of alternatives and the mental capacity that stops the regress.

I propose to take both of these problems seriously. In the next section, I will develop a conception of the process of arbitrary selection in a way that explains its resolution of Buridan cases, avoids the regress, is consistent with the Identity View, and makes policies unnecessary. But first I want to consider one other appealing response to Buridan cases.

4. The appeal to attention

The 6th century Aristotelian commentator Simplicius, discussing a Buridan case, wrote that the protagonist "will choose whatever he first happens on"(1894, 534).[86] In this context, "happens on" can suggest randomness or chance, and also encounter or focus: the man in Al-Ghazali's example can "choose whatever he first happens on," by just taking whichever date his attention turns to first. This thought is articulated in Ullmann-Margalit and Morgenbesser's discussion of cases like his (1977, 774):

> Often enough, or perhaps typically, what occurs…is that you haphazardly focus your *attention* on some one of the available alternatives. Once you do that, however, then—by hypothesis—none of the other alternatives attracts you more, and there is no room for qualms or second thoughts. So, given the absence of either detracting or distorting factors, there is nothing to prevent you from going ahead and grabbing (or doing) that focused-on alternative.

A haphazard focus of attention can be thought of as being in contrast to a normative belief. But it can also be thought of as something to appeal to in defending the Identity View. Al-Ghazali's man can intend to, and so believe he ought to, take whichever date his attention falls on first. That would offer him a way to resolve his Buridan case by forming a normative belief and doing what he believes he ought to: let his attention fall on one of the dates, and take that one. Is this an answer to the Buridan problem?

The appeal to attention is attractive. It seems recognizable as something we in fact do in Buridan cases: when asked to pick, we do often let our attention fall on one date, or card, or bundle of hay, and take that one. The appeal to attention also seems to avoid the

[86] I use Rescher's translation; see Rescher (1959, 145).

problems with the other responses. It avoids both denying the phenomenon of Buridan cases, and conceding that we cannot resolve them. And it offers attention as a mental 'randomizer'. Attention seems to let us delegate our choice to something other than deliberation, without raising the problem of matching acts to outcomes.

Some such use of attention is paradigmatic in resolving Buridan cases, and I think the appeal to it is on to something. But that appeal cannot itself be the solution. Recall the two regress problems for randomization: choosing a randomizing device, and matching random outcomes to available acts. Attention seems to avoid the matching problem. It, so to speak, 'matches itself' to an act, by drawing attention to it. Whether to take the alternative we are attending to or the one we are not attending to does not strike us as a further Buridan case.

But there is still a regress problem. There is a problem of *how* to use attention. Al-Ghazali's man might wonder: do I let my mind wander, or my gaze, or use some more particular, quirky method? And when am I done? What should I count as my attention fully falling on an option? Should I stop when I am looking at one of the dates, or when I'm closer to one, or when I feel a greater inclination toward one? Should I take whichever one my attention falls on next, or, instead, the last one I attended to before I understood that I see no relevant difference between the alternatives? The mere thought of using attention leaves unspecified the way of using it.

The regress problem can also be a problem of *whether* to use attention. This problem is that there can be *other* ways of resolving a Buridan case. Al-Ghazali's man can do a dance, and take whichever date he ends up closer to. Or he can throw sugar at both and take whichever has more sugar on it. Or he can ask someone else to do any of these things. If it occurs to him that he has these options, letting attention fall on one date may look like one possibility among others. Any of them might work, and he might not see any difference between them relevant for belief about which one he ought to take. He is then back in a Buridan case, of a kind it seems we are able to resolve. If we do manage to form an intention in such cases, how do we do it—and how could our doing it involve a normative belief?[87]

The challenge for the Identity View is to account for our intentions in Buridan cases, in which it seems clear that we do not form a normative belief. To do this, it is not enough to account for *most* such cases. The Identity View should leave *no* clear cases of intention without normative belief. But like the appeal to randomization, the appeal to attention at best reduces the number of troubling Buridan cases. It leaves fewer problem

[87] One might still wonder: Doesn't attention save time compared to these other methods? If so, is there still a Buridan case concerning whether to use attention? Here, first, there is no guarantee that it will always save time. Second, even when it does, the person in the Buridan case may not believe it does, and so may still not see a relevant difference between her options. She would then still face a Buridan case. Third, even when she does believe using attention in one particular way would be the fastest, she might find the dancing option slightly more enjoyable, and not see how to choose between the slightly faster and slightly more fun options. She would then be in an incomparability case, which would not advance her very far beyond a Buridan case (see §IV below). For all of these reasons, appealing to the speed of attention does not remove the problem.

cases, but any remaining Buridan cases raise the same problem. So even if it captures something important, the appeal does not show how, in general, a normative belief can be part of our response to such cases.[88]

5. Desiderata for an adequate response

Though none of these initial responses is adequate, considering them brings out more clearly what an adequate response would have to do. To successfully account for intention in Buridan cases, the Identity View must meet

The Exhaustiveness Condition: Account for *all* intention in Buridan cases.

On the Identity View, intention is normative belief. The view cannot leave *any* intention that is clearly held in the absence of, and so cannot itself be, a corresponding normative belief.

None of these initial responses meet the Exhaustiveness Condition. Denying the phenomenon ignores Buridan cases altogether; biting the bullet ignores our intentions in Buridan cases; appeals to randomization or attention cannot account for our resolution of cases in which we must decide how or whether to use these strategies. The last two responses, though more promising, leave some paradigmatic Buridan cases unaccounted for, and so cannot meet the Exhaustiveness Condition either.

In describing these responses, I have described the motivation for a central appeal to arbitrary mental selection, in a way that illustrates some of the desiderata for it. The Identity View must address apparent Buridan cases and their resolution, in a way that captures how we handle them and leaves no Buridan cases unaccounted for. As Rescher's discussion of policies suggests, the Identity View should, ideally, also allow that our intentions in such cases can be ones which we can reasonably defend. *Pace* Spinoza, our handling of Buridan cases does not seem like suicide, idiocy, or madness, and the Identity

[88] Though I have considered four representative responses, I do not claim that they are exhaustive. Tenenbaum (2007 and 2014), for example, suggests an alternative, on which "some inferences can be merely permissible"(2007, 70). Though Tenenbaum's topic is judgments of goodness, one could extend his view to insist that it can be permissible in a Buridan case to simply infer that one ought to take, say, the bale on the left. Perhaps, then, we can directly form an intention in a Buridan case by making such an inference.

Even if Tenenbaum is right about the permissibility of such inferences, this response is again likely to leave some intentional resolutions of Buridan cases unaccounted for. For example, people who do not already held Tenenbaum's view may not allow themselves to make such inferences. They then need another kind of resolution, and may be forced to the one I go on to describe. In any case, I consider the four responses in the text mainly to characterize the challenges a response must meet in order to be successful. As I go on to say, the Identity View can allow for various kinds of resolution of a Buridan case. If it is possible, directly inferring that one ought to take the bale on the left would be one such example. The availability of such an alternative strategy would, I believe, still allow the one I go on to describe, and would sometimes require it.

View should be able to explain or at least allow the rationality of intentions in such cases, at least in principle.

To account for intention in Buridan cases, it seems, the Identity View should:

(1) allow that there are genuine (real or imagined) Buridan cases;
(2) allow that we can form intentions when faced with them;
(3) show how some mental activity allows us to handle such cases;
(4) show how the activity is not intention formation *without* normative belief;
(5) be recognizable as the way we form intentions in Buridan cases;
(6) meet the Exhaustiveness Condition; and
(7) avoid the conclusion that intention and action in such cases must be irrational.

I will try to develop a response that satisfies all of these desiderata.

III. Deciding to Act Non-Intentionally

1. The Basic Picture

Like the appeal to attention, I think it is right to look to a process that is mental but not intentional to explain the resolution of Buridan cases. But it is important to appreciate the philosophical significance of the fact that there is often more than one such process available for intention to rely on. I think the key to a solution is to not insist that we settle on *any* particular non-intentional process. There is no one particular process or strategy that we always intend to use. And when forming an intention in a genuine Buridan case, I think we do not need to intentionally make use of any particular way of going about it. We can simply *decide to act non-intentionally*.

By "non-intentionally," I mean simply "without intention." To act non-intentionally in this sense is to do things we do not (under any description) intend to do. Although "decide" can be used in various ways, I will use it as shorthand for 'form an intention'.[89] To decide to act non-intentionally, in my sense, is to form an intention to act in ways we do not intend to act. Though this way of putting it is paradoxical, the idea is simple. It is not that we can intend to act in some ways, while also not intending to act in those ways. Rather, it is that we can intend to let ourselves act without the guidance of intention.

The category of non-intentional action is familiar. We doodle, we pace while thinking, we hum, and we dodge oncoming objects. When asked why we are doing these things, we would often answer that we do not intend to be doing them. Like the movements of some animals, these actions are performed without intention.[90]

---

[89] Here I follow a common usage on both sides of the debate; a useful list of references is provided in Mintoff (2001, 219n5).

[90] The contrast is sometimes drawn in other ways. Frankfurt (1988b) and Velleman (2000), for example, draw a distinction between action and mere activity, on which, for example, "idly and

Of course, if we *decide* to do one of these things, it is done intentionally. If we do form an intention to hum the Marseillaise, the humming seems no less intended than anything else we do.

But we can also form an intention to do something non-intentionally. We can notice ourselves doodling or pacing intentionally, and take our attention off it to let it continue non-intentionally. Some sports players place enormous importance on letting much of their throwing, dribbling, or running be non-intentional. They know that they do worse when, as one sports psychologist put it, "They start to overthink something that should really be reflexive….It destroys their ability to do what they have been practicing so long."[91] When an athlete goes through a routine to 'turn off' reflective thought and intentional activity, she is, among other things, deciding to act non-intentionally, or let her finely tuned bodily abilities run their course.[92]

---

inattentively" drumming one's fingers on a table (Frankfurt 1988b, 58) or scratching one's head (Velleman 2000, 2) count as activity but not as action. In my usage, the drumming or scratching would be an action, though not one that we intend to perform. I keep this usage partly because it is useful to have an umbrella term, in order to then focus on the presence or absence of intention. At the same time, drumming one's fingers or scratching one's head can be singled out as one thing we do, and it is natural to think of something we do as an action.

On the other hand, not everything that we in *some* sense 'do' may count as an action. Setiya (2007, 23), writes, for example:

> In the course of a typical day, I do a multitude of things: I breathe almost continuously; I blink from time to time; I look at things, pick them up, and put them down; I eat and drink; I read; I listen to music; and I forget something I meant to bring to the office. All of this counts as my behavior—what I *do* in a minimal sense—but not all of it is done intentionally. I *could* breathe and blink intentionally, but mostly I do not.

To breathing and blinking, we could add snoring, which, apart from imitating the sound, we can never do intentionally. I am inclined to disagree with Setiya's view that most breathing and blinking count as something we do, or as behavior. They may be more naturally compared to the circulation of the blood, which hardly counts as 'behavior', or as something we 'do' even in a minimally substantive sense. It is natural to think of all of these as non-intentional bodily processes, rather than as non-intentional actions. But I will not insist on this here, or try to precisely mark the boundary between non-intentional action and non-intentional movement that is not action at all. What is most important in this section is the decision to suspend intentional action and resume it once a Buridan case is resolved. By the same token, if we reserve 'action' for intentional actions, the solution I offer can be redescribed in terms of non-intentional activities, behavior, or processes.

One further terminological clarification will help here. I do not mean "non-intentional" to carry any connotation of being accidental, unwanted, or against one's will, as "unintentional" sometimes suggests. For the contrast, see Aristotle's discussion between the involuntary and the non-voluntary in his *Nicomachean Ethics*, Book III, Chapter I (Aristotle 1999).

[91] The quote is from Dr. Shawn Harvey, quoted in Dreyfus and Kelly (2011, 80).

[92] Here I do not consider whether it is possible to decide to do something non-intentionally, and to act non-intentionally on that intention, or only to decide to *let* oneself act non-intentionally, believing a particular non-intentional action will be the result. I am inclined to think we cannot act non-intentionally 'on' an intention to do something in particular; but this issue does not arise here,

Just as we can form an intention to act non-intentionally, we can follow through intentionally on an already begun and so far non-intentional movement. We can stop at a crosswalk, lost in thought, and then find ourselves halfway across the street with a "Walk" sign in front of us. Whether we find this frightening or natural, we tend to keep walking, thinking of ourselves as intentionally continuing a crossing that we started "on autopilot." We have the capacity to intentionally complete a movement that we find ourselves in the middle of.

In some cases, we can both intend to act non-intentionally, and intentionally continue an already begun movement. When I was in college, I often sat by the river to relax and clear my mind. At first I tried to think about how long would be enough, but I could not settle on any particular length of time. The best solution I could find was to trust my instincts. I decided not to go home until I noticed that I had *already* gotten up and started walking. This usually took about an hour. Whenever the idea of going home occurred to me, I would stay sitting on the grass. But when I noticed that I was already walking, I would continue, trusting my standing up as a sign that I was ready. I had a policy of intentionally walking home when I noticed that I had non-intentionally started walking home.

In this series of events, there are several potential objects of explanation. There is the initial decision, to go home when I notice that I've already started walking. Then there is my standing up and beginning to walk. And finally, there is the complete action of walking home. The Identity View has different implications for each explanation. On the Identity View, the initial decision—if we think of a decision as the formation of an intention—is the formation of a normative belief. On the Identity View, I come to believe: that I should go home when I notice that I've already started walking. That belief can be justified as an application of my policy, and there can be disagreement, concerning it and the policy, about whether the strategy is a good one. One can think of the strategy as a smart way to trust one's instincts and remove distracting thoughts about when to go home—or as poorly thought through, unreliable, less restful than setting a timer, or likely to make my time at the river too long or too short. The belief is directly subject to this kind of evaluation, and is usually itself a response to considerations like these. The standing up and beginning to walk, on the other hand, is like a non-intentional pacing, doodling, humming, or ducking. The Identity View carries no implication that this action requires a normative belief. It says nothing about non-intentional actions at all. Once I do start to walk, and notice what I am doing, I follow through intentionally, acting on my intention to go home when I notice that I have already started walking. Though the Identity View is not directly about intentional action, the explanation of my intentionally walking home would

---

since the intention I am attributing in Buridan cases is simply the intention to act non-intentionally in one way or another. On the other hand, if we cannot act non-intentionally *on* an intention to act non-intentionally, it may be too loose to speak of intending *to* act non-intentionally. The intention may simply be to *let* oneself act non-intentionally, and in that case the phrase "intending to act non-intentionally" must be shorthand for "intending to let oneself act non-intentionally". I leave out this complication in the text, and continue to speak of intending to act non-intentionally.

naturally point to my belief that that is what I should do. It is important to distinguish these different objects of explanation: the intention, the non-intentional action, and the intentional action. They are distinct, even when they are temporally close together.

The resolution of Buridan cases can be explained in the same way. Imagine yourself in a Buridan case, in which you do not see how to justify favoring any one option. Like Al-Ghazali's man, you are hungry and can walk over to one of two dates. Having one date is better than none. You believe this, and believe you should take one. You also believe you need some way to do it. No outside intervention is forthcoming, and neither is a reason to take one over the other. It seems that, on the one hand, nothing other than an action of yours can resolve the case; on the other hand, you see nothing on which to base an intention to take either date, and no way to begin the intentional action of taking one. All you have left is action without intention. So you decide to act non-intentionally as a way of getting one of the dates. Once you have begun to move, of course, the tie is broken. You can see relevant differences between the two dates; one is slightly closer, and you are moving toward it. You can then form the belief that you should take that one. And you can follow through intentionally on the already begun motion. This is what you do, as a means to getting one of the dates.

Such a case is structurally similar to my sitting by the river. At the river, I decided to act non-intentionally. I believed in advance that the action would be standing up and starting to walk home, though I did not know when this would be. In taking a date, you decide to act non-intentionally, and believe in advance that the action will be starting to move toward a date, though you do not know which one. The first case is not a Buridan case, because I did not believe there was *no* relevant difference between standing up earlier and standing up later. But it illustrates a strategy that applies to Buridan cases in general. The strategy is to first form an intention that responds to the available reasons (for going home soon but not thinking too much, or for getting a date) but underspecifies the action to be taken; and then to form an intention to act non-intentionally, in order to determine the action on which we will intentionally follow through. In both kinds of case, we have a belief in advance about what it is that we want to non-intentionally do. We want to begin an action that is, from our perspective, at least as good as any other we could perform. But in order to do that, we let our non-intentional actions decide which one.

To describe non-intentional selection between alternatives, it can be useful to introduce the notion of *picking*. We are sometimes asked to pick a card from a deck, or a cookie from a tray, or a can from a shelf, or a number between 1 and 20. It is less common to be asked to *choose* a card from a deck, or to *decide* on a card. Although these words are used in many ways, the word 'pick' can usefully mark a contrast. When we are asked to pick, we are typically expected not to deliberate. And in agreeing to pick, we agree in part to rely on non-intentional processes. We do not fully understand the nature of those processes. But we normally know we can count on them to make the selection. When we agree to pick a number, we typically allow our non-intentional activity to determine what that number will be.

'Picking' often contrasts with choosing based on a prior evaluation of one

alternative as distinctively worthwhile. Ullmann-Margalit and Morgenbesser (1977) helpfully describe the contrast. As they point out, a single situation, such as buying a hat, can be a 'picking' situation for one person who is indifferent among them, and a 'choosing' situation for another, depending on each person's preferences, expertise, and character. A friend's advice or an advertisement can transform a picking situation into a choosing one or vice versa. And just as there is picking and choosing, there are pickers and choosers. Picking can be more prevalent for those of us who are apathetic, nonchalant, or carefree, and choosing for the pedantic, meticulous, or neurotic. Some of us are unusually indifferent; others treat even minute differences as relevant.

Though the notion of picking can be helpful, it is hardly unequivocal or independently clear. Talk of picking is ambiguous in at least two important ways. First, the word 'pick', as used in ordinary contexts, does not always mark non-intentional action. 'Picking' can be used to describe an intentional and even highly reflective process. A teenager picking a college to go to may consider a complex set of considerations over an extended period of time to come to a paradigmatically reflective intention, followed by a fully intentional action of accepting an offer from a particular college. Here 'pick' can be used interchangeably with 'choose' or 'decide'. Even the phrase 'mere picking', or 'just pick', is not a reliable sign of non-intentional action. After a grueling admissions process, a highschool senior can say in relief: "Now I just pick one and I'm done!" In this case, her 'just picking' is a sophisticated intentional process, in contrast to an even more complex one.

Second, 'picking' can refer solely to a mental process, or include a completed outward action. We can pick a number between 1 and 20 without moving and without saying anything. Picking a cookie, on the other hand, normally involves picking it up. The number and the cookie, of course, do not guarantee one sort of picking or the other. We can blurt out a number without silently picking first, and we can 'pick' the second cookie in the third row on the tray before beginning to reach for it.

It can often be unclear which sort of 'picking' we are doing. Both sorts of ambiguity complicate the range of possibilities. Asked to pick a number, I might say "17" without any prior mental 'picking', or because it is my favorite number and I choose it for that reason whenever I can, or after pathologically painstaking deliberation. All of these can, in various contexts, be described as 'picking'. On the other hand, not every decision to act non-intentionally is naturally described as one in which a person picks. It can seem strained to say that, when sitting by the river, I 'picked' a time to go home. If the non-intentional action of Al-Ghazali's man were to destroy one of his dates, it might strike us as unnatural to say that he 'picks' the other one.

Ordinary usage makes the notion of picking both useful and potentially misleading. Talk of picking itself calls for further specification. So with these complications in mind, I will occasionally use the word 'picking', in a sense that covers both a purely mental picking of a number, and a reaching for a particular cookie. In resolving a typical Buridan case, we can say, we 'pick' in the following sense: we act non-intentionally to determine

which alternative we will take.[93]

2. Recognizability

Is this really recognizable as the way we resolve Buridan cases? The most skeptical version of this question doubts that we can take this strategy at all, or at least with any consistent success. Try not to think about pink elephants. You will usually fail. Or, at best, you will succeed without the ease with which we pick a card or a soup can. So why think we can decide to act non-intentionally with any more success? A baseball player might need an elaborate pre-game routine to enter a mode in which he can take his swings without intention. Can we turn off intention so easily in Buridan cases?

A similar doubt can begin with introspection. When we consider the way we resolve Buridan cases, the decision to act non-intentionally can seem to build in too much cognitive structure. When grabbing a soup can off a supermarket shelf, do we really decide to do it non-intentionally, then start to do it, notice ourselves taking one, and intentionally follow through? One wants to say: don't we just take one and buy it? Even if the decision to act non-intentionally is possible, and even if it is effective, it may involve more than we can recognize as what we do in these cases.

My afternoons by the river are enough to establish the possibility of success in the intention to act non-intentionally. The more difficult issues here are not whether we can *ever* succeed, but reliability, effort, and overall recognizability. With respect to reliability and effort, trying not to think about pink elephants offers a helpful *dis*analogy. When you are told not to think about pink elephants, you are set up to fail. The intention not to think about them has a content which, when called to mind explicitly, frustrates the intention. To remember the intention is to fail to carry it out. So although we can succeed in distracting ourselves through indirect means, the attempt is essentially effortful and unreliable. When we decide to act non-intentionally in a Buridan case, on the other hand, we are normally set up to succeed. The attempt leaves our attention directed toward the options we see as most worthy of choice, and we usually have a desire for each of them. The rest of our ongoing activity tends to upset the delicate balance of the Buridan case, while our attention and desire are often pulled to settle on one option and dwell on it. At the same time, we have an enormous wealth of experience with Buridan cases, especially when it comes to the objects of appetites like hunger or thirst. We have picked one cookie out of many, more times than

[93] Here my usage differs from Ullmann-Margalit and Morgenbesser's (1977), in two ways. First, rather than treating 'picking' as the central notion throughout, I insist that it calls for further specification in terms of other, more fundamental notions. This paper can be seen as offering a conception of the relevant sort of picking. Second, in considering responses to a picking situation, they distinguish "direct" extrication—that is, by picking—from turning the situation into one that is no longer a picking situation. In my usage, the contrast between non-intentionally reaching for a cookie, and non-intentionally moving in a way that leaves one cookie much closer and therefore obviously privileged, becomes irrelevant.

For an earlier treatment of a related notion of "non-rational choice," and of "choosing to choose at random," see McAdam (1965).

we can count, and resolved many related Buridan cases. We have a well-developed repertoire of habits of non-intentional action for these and related cases, and we can usually just let them take their course. It is no surprise that a decision to do this can be reliably effective.

This point about habits also helps with the recognizability of the strategy. The challenge for the Identity View is not to account for the existence of, or even the possibility of action in, Buridan cases. It is to account for *intention* in Buridan cases. Recurring Buridan cases, such as picking a coin, or a card, or a can off a shelf, are cases we tend to resolve quickly and habitually, without first forming an intention at all. It is only when we have to form an intention that the strategy applies. So the strategy does involve much more than usually happens in a Buridan case. It is what we do, only when the resolution of the case involves an intention.

To see this contrast, imagine the indefinite number of slightly different possible ways of raising your arm. When you raise your arm, there might be many such alternative raisings between which you see no relevant difference. And when we raise an arm, we usually do it without thinking or deciding how to do it at all. But when we do have a more particular intention—to mimic someone's arm raising, or to keep our wrists straight—that intention brings greater cognitive structure to the action. This, not the ordinary raising, is the analogue to the intentional resolution of a Buridan case. When, for example, we get so far as to ask ourselves which soup can to take, we find no easy way to choose a particular can. It is at this point that we usually say: "I just have to pick one." This recognition is, I think, an expression of the need for a non-intentional action, a 'mere picking' in contrast to decision. It is here that I think we can recognize ourselves taking the strategy I described. We decide to 'just pick' and to then 'go with that one'.

3. Exhaustiveness

Recognizability does not yet show that a view meets the Exhaustiveness Condition. Even if we sometimes intend to act non-intentionally, does this strategy allow intellectualism to account for *all* intentional resolutions of Buridan cases?

There are many ways for a Buridan case to become a non-Buridan case. A passerby can eat one of the two dates, or bring one over, or point out that one is moldy. Or instead, we can decide to look more closely until we find a relevant difference. Or instead, a particular way of using attention, such as looking around until our eyes look directly at one date, might occur to us, without any other apparently equally good strategy coming to mind. Or instead, we can convince ourselves that the best alternative is to avoid sugar altogether, and make an omelet instead. In thinking philosophically about these cases, we do not need to rule out that any of these are possible or even common. As before, the Identity View need not insist that any particular strategy is the only one we ever take—even the strategy of deciding to act non-intentionally.

To account for our responses to Buridan cases, the Identity View must leave no case in which we form an intention without being able to form a normative belief. This was

the initial power of Buridan cases. They seem to be not just foot-stomping insistence, but a set of compelling examples in which there is *no* basis for a normative belief, and yet success in forming an intention. Buridan cases seem to be cases of deliberative failure, in which an inability to see a single privileged alternative forces us to select one without believing it is what we ought to do.

If "select" means "intend," we are never forced in this way. It is not true that, if the donkey and the man could act only as reason commands, they would starve. The intention to act non-intentionally offers a way to resolve Buridan cases through normative belief. Sometimes, the route to the normative belief can run through an entirely different explicit strategy, such as turning in place or flipping a coin. But even when we have no such strategy in place, and none comes to mind, we can turn to non-intentional action, believing that this is what we should do.

Here, deciding to act non-intentionally is not simply one recognizable strategy among others. When we see no way to resolve a Buridan case, there is a kind of deliberative pressure to decide to act non-intentionally. Faced with two identical dates, you cannot pursue a resolution except through some action of yours. But if you are truly stuck—if you are deliberating, and deliberation is at a standstill—you see no way to form an intention that will resolve the case. You see no way to form an intention; but to resolve the case, you will have to act. If you combine these two thoughts, you will have one: to resolve the case, you will have to act without intention. Once you see this, of course, you can intend to do just that. The decision to act non-intentionally is thus a natural last resort; seeing that we see no way to intentionally resolve a Buridan case itself leads us to turn to non-intentional action. There is still no guarantee of success; but there is a deliberative route to a promising and familiar strategy. The strategy is available, and in one way deliberatively favored, in every intentionally resolved Buridan case. In this way, intellectualism is left with no intentionally resolved Buridan cases in which there is clearly no way to form a normative belief.[94]

This response on behalf of the Identity View might still seem not to explain what it needs to explain to *account* for our responses to Buridan cases. It can seem that we wanted a *contrastive* explanation—one that explains why someone intends to take, and intentionally takes, one course of action rather than a different one. The Identity View should be able to offer some explanation in terms of normative belief, but on the view I described, the selection seems essentially brute. There seems to be no normative belief, and no reason for one, corresponding to the intention to take one particular course of action rather than another. A contrastive explanation in normative terms seems lacking.

---

[94] Note that, unlike randomization, this deliberative route does not give rise to a problematic regress. Someone can face a choice between deciding to act non-intentionally and deciding to turn in place, and not see a relevant difference between these. But if she is unable to form an intention with respect to them, she will normally be led to act non-intentionally in picking a meta-strategy. Unlike a coin toss, this strategy does not itself require her to resolve a further Buridan case. She herself might, in a pathological case, regress indefinitely; but in that case there is no new problem for intellectualism, and no theoretical view can help her.

To see how the explanation is both normative and contrastive, consider a simpler example we can call *Coin Toss*. You and a friend are deciding which of you will organize an upcoming reception. You decide that one of the two of you should do it, you are equally capable, you both equally strongly want (not) to do it, and you see no good principled way to choose, at least not within reasonable time constraints. So you decide to flip a coin and do whatever it says. Coins are your favorite randomizing device, and you are partial to tails and your friend to heads, so there is no Buridan problem. The coin falls tails and you are the organizer. Why is it you and not your friend? Though the organizing itself is intentional and highly cognitively structured, the contrastive explanation can seem completely brute. It seems all in the coin toss. So can we give a contrastive explanation of your organizing rather than your friend's, under the guise of the good?

It is not an accident that neither defenders nor attackers of the guise of the good are worried about this kind of case. In Coin Toss, the course of action was thought through and decided on in advance. Although it left the organizer to be determined by the coin toss, the contrastive explanation essentially has to appeal to the way your normative beliefs set up the situation to depend on the coin. Normative belief in general concerns and responds to features of the environment that are not under our control. So although the coin toss that 'settled' the issue has no normative belief to explain its outcome, what in fact settles the issue is your intention to follow through on the results of the coin toss: your intention, in advance, to accept whatever the result is, and your resulting intention, after the toss, to follow through on tails. That is why you are the organizer and not your friend. Without those intentions, the toss would settle nothing.

When we form an intention in Buridan cases, we treat our own non-intentional action as in one way like a coin toss. We use it as an arbitrary selection procedure.[95] And it is true that my explanation of *which* non-intentional action I take does not appeal to normative belief. But each intention is explained as a normative belief. The original intention to act non-intentionally is a normative belief that one should; and when a person has a contrastive intention to take one course of action rather than another, that intention is explained as a normative belief, and justified by the non-intentional action. So there is a contrastive explanation of every intention in normative terms.

There is another, deeper concern about explanation: the entire attributed structure can still seem unexplained. How do we manage to take our own intentions out of the picture and get ourselves to act non-intentionally? How does the non-intentional action itself happen? The view I describe can seem to point to an unusual process without telling us how it works.

---

[95] Compare Ullmann-Margalit and Morgenbesser (1977, 773): "We are in a sense transformed into a chance device." An implication of this 'coin toss' view of one's action is that, in resolving Buridan cases, we are to some extent dissociated from our own activity. I take a kind of external perspective on my own standing up at the river, and on my beginning to reach for one of the many soup cans on a shelf. As in the previous chapter and the next one, one can see how my account of each of the central counterexamples attributes a kind of disunity or dissociation, and thus a limitation on the person's responsibility for the action. I leave out discussion of dissociation and responsibility in the text.

These questions again help to pinpoint the kind of explanation the Identity View needs to give. How we act non-intentionally, and how we succeed in an intention to act non-intentionally, are interesting and important questions in their own right. But they are not what is at issue here. The question at issue is how intentions in Buridan cases can be normative beliefs, given that normative beliefs privilege one alternative in a way that seems impossible in Buridan cases. The answer is that intentions, when formed in Buridan cases, do privilege one alternative—either the newly discovered alternative of acting non-intentionally, or, in the follow-through, the alternative of continuing what non-intentional action started. The needed explanation is unobvious, but it is neither a general conception of non-intentional action nor a conception of the effects of intending to act non-intentionally. The explanation is that, when deciding what to do in a Buridan case, we can decide to act non-intentionally; that this is the only way we can form an intention that resolves a Buridan case; and that we already understand this. We see that we 'just have to pick', and that is what we do. This is the substantive insight that supports the Identity View.

4. Rationality

We have already begun to see why it is justifiable to decide to act non-intentionally, to intend to follow through, and to in fact intentionally follow through with the alternative we non-intentionally select. Although Bratman (1987) sees Buridan cases as a central challenge to views like the Identity View, he himself offers a source of support for the Identity View in this respect. As he puts it: "My desire-belief reasons in favor of taking route 101 to San Francisco may seem on reflection equal in weight to those in favor of route 280. Still I must decide."(1987, 23). Why must I decide? Because I want to get to San Francisco. This justification is pressing; in Bratman's words, "He really must settle the issue and get on with his life"(2007, 148). And it applies even before one physically hits the fork in the road at which the paths diverge. Even before we take one route, we often need a plan. We "have limited resources for use in attending to problems, deliberating about options, determining likely consequences, performing relevant calculations, and so on"(1987, 10). Buridan cases themselves give us reason to form such plans. "The need for decision in the face of equidesirability, when tied to our needs for coordination, provides independent pressure for being a planning agent"(1987, 11-12). We want to achieve the larger goals that give rise to our Buridan cases, and, given our limited resources and the need to coordinate with other plans and other people, we often need to decide on an alternative in advance. We thus have good reason to seek *some* intentional resolution of Buridan cases.

Instrumental reasoning offers the next piece of justification. We want (or intend) to resolve the Buridan case. When we are stuck in such a case, intending to act non-intentionally is our only way to go about reaching a resolution. So it is rational to intend to act non-intentionally, as a means to resolving the Buridan case.

Once the Buridan case is resolved, of course, justification proceeds as usual. Starting down route 101 gives us obvious reasons to continue down it rather than turn around and take route 280. At that point, route 280 would be slower and more cumbersome.

The more difficult step to justify is the mental activity before the fork in the road. What is the justification for settling on one option then? Bratman suggests that there might not be one. Using a distinction between judgments "made from a standpoint external to the standpoint of deliberation" and judgments from within it (1987, 51), he writes (1987, 45):

> Suppose I arbitrarily decide to take route 101 rather than route 280, even though at the time of my decision the routes seem to me equally attractive. Once I make this decision, my taking route 101 will be rational from my internal perspective, whereas my taking route 280 will not be, for it will be inadmissible. But from the external perspective each option may well remain equally desirable—until I begin driving toward route 101 and away from route 280.

Bratman treats beginning to *drive* toward route 101 as the point at which, from outside deliberation, one alternative begins to be more desirable than another. But a mental act can play the role of tie-breaker just as well. As Bratman emphasizes, we engage in complex coordination of our plans with each other and with those of other people. Even before we take route 101, and even before we tell someone we will, we can find that our minds have just wandered and we have thought through some of the details of the route 101 path. This can play the role of the non-intentional action that resolves the Buridan case, just as driving can. As before, a person might not have any theoretical understanding of what she is doing when she does this. She just has to be able to do it, and to notice that she has. The non-intentional action then provides a consideration that makes one alternative slightly but clearly more worth taking than the other. That consideration justifies forming the intention to take it.

Bratman also helps to explain the final piece of justification, for following through on the already formed intention. As he puts it, "My intention resists reconsideration: it has a characteristic *stability* or *inertia*…. Lacking new considerations I will normally simply retain my intention up to the time of action. Retention of my prior intention and nonreconsideration is, so to speak, the default option"(1987, 16-17). It is the default for good reason: "given our limits and the importance of plans in reliably extending the influence of present deliberation to future action we may expect that reasonable habits of (non)reconsideration will involve a tendency not to reconsider a prior plan except when faced with some problem for that plan"(1987, 66-67). I am inclined to think that a good enough reason to reconsider can be more slight than a *problem* for the plan, especially in a Buridan case. If not much hangs on it, we can justifiably switch to route 280 just for the fun of switching, or to satisfy an urge for change, or to have an example to consider to see

whether it undermines the Identity View.[96] But Bratman's larger point still holds. Limited intellectual capacity and demands of coordination create a justifiable default habit of nonreconsideration.

At this point another suspicion may arise. Intending to act non-intentionally is simply intending to let one's activity continue without intention. But if no other intention is available, it seems one's activity will continue without intention, whether one intends it to or not. So one may wonder: what is the justification for *intending* it?

There are several. First, a person may not see that she can or should act non-intentionally to resolve a Buridan case. She may then intend *not* to let her activity continue without intention. This can frustrate her intention to take one of the alternative actions in the Buridan case. That acting non-intentionally can be the only way to *succeed* in resolving a Buridan case is a justification for intending it, not a sign that the intention is superfluous. In other words, it is not true that one's activity will continue without intention whether one intends it to or not. One may, in some cases, remain 'stuck' or paralyzed, or give up on all the alternatives and move on to a different activity.

Second, though we often find it easy to non-intentionally select an alternative in a Buridan case, proceeding without intention can in some cases be difficult. If we are tempted to keep looking for reasons for another intention, we may need to take further means, such as distracting ourselves with a mantra, to ensure that we act non-intentionally. Intending to act non-intentionally can help us resist the temptation to deliberate.

Third, in some cases, intending to act non-intentionally can lead us to take further means to influence our own non-intentional action. If we intend to non-intentionally select a date to eat, we may also, intentionally, take actions that we believe are likely to speed up the non-intentional selection. We may, for example, intentionally think about how much we like dates, if we know this tends to get us to select more quickly. For all of these reasons, the intention to act non-intentionally is far from superfluous.

The justification I have given here is more complex than the one a person herself will usually have to give in a Buridan case. But it is a way of spelling out the justification implicit in the person's own thought, if she has one, about why she decides to act non-intentionally. The justification is already implicit when someone thinks: "Well, it doesn't matter which one. And I have to decide. So I guess I just have to pick one." Of course, to form the intention to act non-intentionally, we do not have to have an explicit thought about why we must. Al-Ghazali's man, or a shopper in a soup aisle, can just think: "I should just pick one and go with it. Mmmm… okay, this one." She does not need to have a theoretical grasp of, or a further justifying thought about, what happens in her mind when she thinks: "Okay, this one." But if I am right, then, *if* she is deliberating, she usually has a thought like: "I should just pick one and go with it." That would be a rational thought to have, and to act on.

---

[96] When these reasons to reconsider are slight, they may, of course, be clearly too slight to warrant reconsideration. In other cases, it may be unclear to someone whether it is worth reconsidering. In that case she may find herself unable to compare her options, in the way I go on to discuss in the concluding section below.

5. The Desiderata and the Initial Responses

The possibility of deciding to act non-intentionally offers a way for the Identity View to account for intention in Buridan cases that meets all of the desiderata I described in §III. It allows that there are genuine Buridan cases, and that we can form intentions and act intentionally when faced with them (desiderata 1 and 2). It describes a mental activity that allows us to do this (desideratum 3). The mental activity that resolves the Buridan case falls short of intention formation without normative belief (desideratum 4), because it is a combination of normative beliefs with a non-intentional process that itself involves no intention formation at all. The activity is, I argued, recognizable as the way we form intentions in Buridan cases (desideratum 5). As I also argued, the strategy meets the Exhaustiveness Condition (desideratum 6). And there is no implication that intention and action in Buridan cases must be irrational (desideratum 7), since the intention of taking one of the preferred alternatives, the intention to take a particular one, and the actual taking of it can all be given a principled justification.

The decision to act non-intentionally also helps to explain the appeal of all four initial responses considered in the previous section. When it comes to denying the phenomenon of Buridan cases (response 1), it would still be too quick to say that there is always some relevant difference in the object, no matter how small. There is no obvious impossibility in the idea of two exactly identical objects in a symmetrical room. But there is always, so to speak, some difference in the subject. Living beings move, look around, explore their world, do one thing while hesitating about another, let our thoughts wander across various alternatives, and otherwise engage in activity that will at some point break such a symmetry. We can rely on ourselves to do this.[97] Beings like us may not be guaranteed to find a difference between equally good options, but we are usually guaranteed to create one. We can expect never to run into insoluble Buridan cases.[98]

[97] I do not mean to suggest that our relation to our non-intentional action is essentially passive. We can rely on ourselves to act non-intentionally, and we can also put ourselves in conditions in which it is likely to happen more quickly or effectively. But the possibility of being caught in a Buridan case is already unlikely, for reasons partly like the ones the deniers of the phenomenon suggest. As Leibniz (1952, §46) put it: "Innumerable great and small movements, internal and external, co-operate with us, for the most part unperceived by us."

[98] I think there still is the possibility that, due to some contingent obstacle, a living being may be unable to pursue an available solution. There might be pathological cases, or intelligent non-humans, who are unable to decide to act non-intentionally and follow through intentionally, or unable to recognize that they can do this. They may have an impairment in reasoning, or an intense localized anxiety, or an obsession with reversing any initiated action to return to the equilibrium of the Buridan case. These failures do not threaten the Identity View, since they leave no intention (or no intentional action) to explain. And I think they are compatible with the intuition that no Buridan case is insoluble. As in arithmetic, every problem can be soluble even if we occasionally fail to reach the solution. Moreover, since, when we have the intuition that no Buridan case is insoluble,

Though this does not show that there are no genuine Buridan cases to be resolved, it explains some of the appeal of that idea.

On the other hand, the Identity View can also say why there seems to be something to Montaigne, Leibniz, and Spinoza's insistence that the donkey would starve (response 2). Buridan cases frustrate deliberation. When deliberating about them, we see no way to directly resolve one by forming an intention to take one of the candidate options. We avoid paralysis only by taking a different kind of option from any of the alternatives we seemed to be faced with. When the alternatives remain unchanged, intentional resolution of a Buridan case cannot proceed on its own, without recourse to non-intentional action. As earlier philosophers might have put it, the rational will must rely on the animal will to achieve its purpose.[99]

---

we tend to ignore the possibility of such special cases, we can normally leave them out in accounting for the intuition. This is why I do not pay much attention to the possibility of real paralysis when considering Buridan cases in the text.

Still, it is worth noting that radically different cases of non-human intelligence may raise other problems. For example, my solution may not easily apply to a divine will. An omniscient, omnipotent God may not have anything like an animal nature that can act non-intentionally, and whose result He would not know in advance. Here a different kind of response might be called for. One difference may be that God does not have intentions, or, in general, volitional attitudes whose execution requires effort or, in the case of future-directed intention, must be put off until later. If temporal concepts apply to His will at all, God may always be able to act immediately and effortlessly. There may also be a special theological problem about God's resolution of Buridan cases, analogous to the problem of whether an omnipotent being can create a weight He cannot lift, or a problem He cannot solve.

Can God be in Buridan cases, and can He resolve them? On this divine variant, see Ullmann-Margalit and Morgenbesser (1977, 774n19), and Strickland (2006). The topic is a large one, connected to problems about divine freedom more generally. Some might find it to be beyond our understanding, or, instead, to show the incoherence of the notion of an omniscient, omnipotent God. But I am inclined to the simpler view that God, the angels, and many other divine beings have the capacity to act non-intentionally, without having an animal nature. *How* they do it might be, like divine causation in general, beyond our understanding. But there is so far no obstacle to thinking that, if divine beings do form intentions in Buridan cases, they can do it through non-intentional action.

[99] Buridan cases thus still present a problem for the guise of the good when the view is applied to all action, rather than only to intentional action or intention. This problem offers one reason to resist the slide from thinking that beings like us can act under the guise of the good to thinking that we always do. Boyle and Lavin (2010, 193) write, for example:

> A subject possesses the power of practical reason only if his reflection on what to do is in general determinative of what he actually does do…. A rational agent must act under the guise of the good in the sense that he must in general pursue ends in virtue of taking there to be something good about those ends.

If "in general" means "in principle," rather than "usually," this thought is committed to denying the possibility of *any* action that is not under the guise of the good, at least Boyle and Lavin's weaker sense of taking there to be something good (see Chapter 1 for discussion). We should not rule out

There is thus an important grain of truth in the appeal to randomization (response 3). In Buridan cases, the person attempting to form an intention must delegate her selection to a process that, from her point of view, is entirely arbitrary. We use our capacity for non-intentional action as we would use a coin toss, though without the regress of assigning values to the outcomes. The difference is that the 'randomizing device' is our own action. So unlike the appeal to randomization, the account I described also offers a way to do justice to Al-Ghazali's view that the will has 'a quality the nature of which is to differentiate between two similar things'. It is our own activity that we use to arbitrarily break the paralyzing tie, though not directly by forming an intention.

We can also begin to see why it is natural to think about attention in the context of Buridan cases (response 4). Though I have not offered a general conception of non-intentional action, it is hard to imagine how non-intentional action could proceed without attention in a central role. Wandering thoughts or eyes tend to settle on one possibility, just as desire or attention itself turns to the possibility of acting right now. Processes like these are naturally seen as playing a key role in non-intentional action at least much of the time, though not always in one particular way. And attention comes close to offering the unique, mental, but non-intentional 'randomizing' device that non-intentional action makes available to us.

Bratman (1999, 220) wrote that if we think of intention as involving a strong evaluative endorsement, then "we have our Buridan problem. It seems that I can just decide on which bookstore to go to, while continuing to see each option as equally desirable." I have tried to explain why this is not true, partly by distinguishing different objects of explanation: in this case, the intention to go to a bookstore, the intention to act non-intentionally and go to whichever one I select, the non-intentional action in pursuit of a particular bookstore, the intention to go to that one, and the intentional action of going to it. I can continue to believe that, before I decided, it would have been just as desirable to go to a different bookstore. But I cannot believe it would be just as desirable to go to a different bookstore now—to, for example, change my mind and walk or direct my plans in a different direction. As I reach for one soup can and put it into my cart, I can believe that another would have been just as good, but not that continuing with this one is no more desirable than putting it back and taking a different one. What I believe I ought to do is, throughout, what I intend to do: pick a can and buy it, or pick a bookstore and go to the

---

the possibility of non-intentional action in beings who act intentionally. My treatment is meant to leave open, and rely on, the possibility that even we act as animals do much of the time.

There is also still the possibility of applying the account in this section to non-intentional action, if the action itself relies on non-intentional movements that are not actions. I do not argue systematically against this possibility here, though I am somewhat skeptical. The action could not be the formation of an *intention* to move non-intentionally, and it is unclear how it would proceed if not by explicitly using an intentional description, at least of the simple kind we use when we decide to "just pick." It might also have to bite the bullet (initial response 2 above) more often, since, when non-intentional movements are not available to solve the problem, unintentional action would not be available either, and paralysis would follow. I leave these issues aside here, since my focus is on intention.

bookstore I picked.

This is how, even in Buridan cases, we intend under the guise of the good. Like *akrasia*, Buridan cases provide no knockdown counterexample to the Identity View.

IV. Existential Choice and Incomparability

Buridan cases are one kind of situation in which we do not see how to reason our way to a choice between alternatives. They are the sharpest and simplest case, in which we believe the alternatives to have no relevant differences between them at all. But there are many other hard cases, in which we do see a relevant difference but still do not see how to form a normative belief. Some of these differences can be so important to us that we cannot even begin to see how to compare them. These cases are both important to us personally, and a related source of resistance to the Identity View. It is worth saying how a treatment of Buridan cases can address them. [100]

Sartre (1957, 24-25) gives one classic example. A young man who came to see him during World War II had to choose between joining the resistance and staying to care for his mother:

> The boy was faced with the choice of leaving for England and joining the Free French Forces—that is, leaving his mother behind—or remaining with his mother and helping her to carry on. He was fully aware that the woman lived only for him and that his going-off—and perhaps his death—would plunge her into despair. He was also aware that every act that he did for his mother's sake was a sure thing, in the sense that it was helping her to carry on, whereas every effort he made toward going off and fighting was an uncertain move which might run aground and prove completely useless; for example, on his way to England he might, while passing through Spain, be detained indefinitely in a Spanish camp; he might reach England or Algiers and be stuck in an office at a desk job. As a result, he was faced with two very different kinds of action: one, concrete, immediate, but concerning only one individual; the other concerned an incomparably vaster group, a national collectivity, but for that reason was dubious, and might be interrupted en route. And, at the same time, he was wavering between two kinds of ethics. On the one hand, an ethics of sympathy, of personal devotion; on the other, a broader ethics, but one whose efficacy was more dubious. He had to choose between the two.

The young man was not choosing between two philosophical theories; he was choosing a

[100] I leave open the possibility of other, closely related cases, in which we doo see relevant differences but still believe the alternatives to be equally worth of choice. To put it colloquially, the differences would cancel each other out. These are not, strictly speaking, Buridan cases. But since it is relatively obvious that they can be handled in the same way, I focus on the less obvious extension to incomparability.

way of life, along with the values it embodied. His choice is an example of what can be called existential choice: a choice that shapes one's identity, without being based on a conclusion about justification. An existential choice defines us, without our being able to say why one alternative is better, or more worthy of choice, or more justifiable as what we ought to choose, than the others. It is an especially deep kind of choice that does not seem made under the guise of the good. The young man may choose who he is or will be without any confidence that a very different path would not be just as good.[101]

There is no fine line between a choice that shapes one's identity and a choice that does not. The young man can make a similar choice on a smaller scale when he spends an evening helping a friend leave to fight in the Free French Forces at some personal risk to himself and his mother. For the Identity View, the key feature of these choices is not their depth, but their apparent lack of justification. What is crucial is that the young man is choosing between "two very different kinds of action," without a sense of how to weigh or evaluate the differences. Glaring as they are, the differences between caring for his mother and joining the resistance seem to him impossible to compare in deliberation. Existential choices are especially important, though not clearly demarcated, examples of incomparability.

Various values, or various bearers of value, are often thought to be "incommensurable": fairness and pleasure, respect and love, Michelangelo and Klee. But that word is used in at least two different senses. Items are incommensurable in one sense when they cannot be measured by a single scale of units of value. They are incommensurable in a second sense when there is no true comparative statement about them with respect to some value. These two senses are at least partly independent. It might be clearly true that Michelangelo was more talented than Rubens, though no precise measurement can be made to say by how much. But, some think, it might not be true that Michelangelo is either more talented than Klee, or less talented, or equally talented, or even roughly equally talented. There might be no comparison to be made here at all.

Following Chang (1997a), we can call the second of these senses "incomparability."[102] It is a substantive and controversial issue whether there is such

---

[101] My usage here is similar to Allan Gibbard's: "Such a commitment we can call *existential*: it is a choice of what kind of person to be, in a fundamental way, come what might, which the chooser does not take to be dictated by considerations of rationality….On our ordinary way of thinking, it seems possible for a commitment to be existential in this sense"(1990, 168). Gibbard describes Sartre (1957) as his inspiration (168n11), though Sartre, and Gibbard following him, adds a thought of "choosing for all mankind" which I do not consider here. Ullmann-Margalit and Morgenbesser (1977, 783-5) similarly mention "the Existentialist notion of the *absurd*" in their concluding discussion of what they call "deeper-level picking." There they write: "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….It just may be that, whether to our delight or to our dismay, it is picking rather than choosing that underlies the very core of our being what we are"(783-785). See also Korsgaard (1996), esp. Chapters 3-4, for a related treatment of what she calls "practical identity"; Bratman (1999b) helpfully ties Korsgaard's view to a discussion of existential choice.

[102] I also follow Chang in thinking of incomparability as the less explored and more significant notion, and in thinking that precise measurement by a single unit of value is not essential to most

incomparability. But it is not at all controversial that alternatives can *seem* to us to be incomparable. It is not controversial that some people believe in incomparability, and make individual judgments of incomparability. And even those of us who deny incomparability can find ourselves unable to reach a comparative evaluation in some cases. We might believe that there must be a fact of the matter about which of Michelangelo and Klee is more talented, but we do not know enough about art history or aesthetics to reach the right conclusion in the foreseeable future. We then have a case of what we might call *subjective incomparability*. Items are subjectively incomparable for a person with respect to a value when she is unable to reach a comparative evaluation of them with respect to that value. She is unable to compare, for example, Michelangelo and Klee with respect to talent. She can be unable to do this, whether or not there is a fact of the matter.

For Sartre's young man, caring for his mother and joining the resistance are subjectively incomparable with respect to justification for belief about what he ought to do.[103] Since this kind of case is what is crucial for the guise of the good, I will call it an *incomparability case*. An incomparability case is a case of subjective incomparability for a person with respect to justification for belief about what that person ought to do. The existence of such cases is independent of whether there is incomparability in the objective sense. And it raises a general problem for the guise of the good, and in particular for the Identity View. In incomparability cases, it seems, we can form intentions even when we are unable to reach a normative belief. And so, it seems, the intention cannot itself be the belief. This is a variant of what I earlier called the Buridan Argument against the Identity View (§I).

Incomparability cases are not Buridan cases. In Buridan cases, a person believes there is *no* difference between the alternatives relevant for belief about which one she ought to take. She then normally believes the alternatives are equally matched with respect to justification for belief about what she ought to do. That is a substantive comparative evaluation, though not in favor of either alternative. In an incomparability case, she has no

---

views. This structures what I say in the text without being explicit there. Some of the recent work on 'incommensurability' in both senses is collected in Chang (1997), with a highly substantive and especially useful introduction by Chang.

[103] I say "with respect to justification for belief about what he ought to do," rather than "with respect to what he ought to do," because in the latter, it may not be possible to reach a judgment of equal value. It is hard to understand what it would be to be believe that two options are equally what one ought to do. Is it that one believes one ought to do both? That is impossible, and so, most of us would think, not something most of us could believe. Does one believe one is in a moral blind alley, in which, whatever one does, one will fail to do something one ought to do? That does not seem to match the conclusion one makes in these cases. Does one believe neither is what one ought to do? That may be a more plausible line to take, but it belongs later in the deliberative process. At this point, the deliberator may not yet have concluded that she is not required to take either option. She may be trying to form a normative belief, and finding multiple options equally attractive. More generally, "what one ought to do" is not easily thought of as a value with respect to which various items could be compared. Something's being "more" what one ought to do than something else is no easier to understand. I avoid these issues by talking of justification for normative belief, rather than directly about the predicate applied in the belief itself.

such belief. The young man is not ready to conclude that he would be equally justified in believing he ought to care for his mother or in believing he ought to join the Free French Forces. He just does not know what to think. This is a different and widespread kind of case. But I think it can be handled in the same way as Buridan cases. I will try to say why.

Incomparability cases overlap with Buridan cases, in the sense that it can be unclear which one we are dealing with. In my earlier example of sitting by the river, I might have had reasons to leave earlier and later than I did, which I was unable to compare. It is natural to think of this case as an incomparability case, in which I cannot come to a conclusion about whether getting home earlier or getting more rest at the river do more to justify going home at a particular time. I decided to let non-intentional movement take over, partly because I could not come to a confident conclusion about when the considerations were equally balanced. In other cases, one can be unsure whether the differences between alternatives are relevant. When choosing between two bookstores, either of which is likely to have the book one wants, one might know that one has slightly better air conditioning and the other has a slightly more pleasant décor. One might be unsure whether the differences are relevant, or think they potentially might be but decide to treat them as insignificant. These choices shade off into cases like the variant on the two dates, in which one is marked "8371" and the other "8713." The two alternatives are clearly different, since the numbers are different. But it is probably clear to the chooser that the difference is most likely completely irrelevant.

As these intermediate cases bring out, treating a situation as a Buridan case can itself be a matter of choice. As Ullmann-Margalit and Morgenbesser (1977) put it, we can choose to just pick—that is, choose to be indifferent between the alternatives. We might, for example, believe there is an optimal rain hat to take, but be in a hurry and pick one without thinking through the alternatives. Or, when choosing where to stay on vacation, we can be at a loss, and choose to pick. We can treat our alternatives as equally good, even when we believe there are relevant differences between them. Just as I did at the river, we can choose not to deliberate, and to let non-intentional action make the choice for us.[104]

[104] Ullmann-Margalit and Morgenbesser (1977) helpfully describe the various combinations of, as they put it, picking (where we are indifferent with respect to the alternatives) and choosing. What I describe here is choosing to pick. We can also pick to choose, when we can have either one alternative, A, or a choice between two other alternatives, B and C, and are indifferent to whether we have A or the choice. Typically, this is because A and B are identical and C is clearly worse. We can also pick which choice to face: say, between A and B or between C and D. A single situation, such as selecting a tie, can be a picking situation for one person and a choosing situation for another, depending on preferences, expertise, and character. A friend's advice or an advertisement can transform a picking situation into a choosing one or vice versa. And just as there is picking and choosing, there are pickers and choosers. Picking can be more prevalent for those of us who are apathetic, nonchalant, or carefree, and choosing for the pedantic, meticulous, or neurotic.

The deeper issue which Ullmann-Margalit and Morgenbesser do not consider in detail is whether picking is itself an act. This issue, I think, is partly concealed by a limitation of their "picking" terminology, since "picking" is itself ambiguous between a mental event—picking a number, picking which can one will take—and the observable movement of taking what one has

In other cases, we can be forced to treat a situation as a Buridan case by a lack of knowledge. We can pick a card from a deck, knowing it matters which card we get, just as a game show contestant picks one of three doors, knowing one of them reveals a car, one a cash prize, and one a rubber chicken. Here we know there are directly relevant differences between the alternatives, but we have no way of using these differences to come to a belief about which alternative we ought to take. We can, of course, try again to find a way to make a comparison—just as, in an apparent Buridan case, we can check to see whether we can find a relevant difference between the alternatives. But once we establish that we are in an incomparability case, we then usually have no alternative but to treat the situation as a Buridan case.

This is an essential feature of incomparability cases. In these cases, we are unable to compare the alternatives. The young man comes to Sartre because he does not see how to choose. On that day, of course, he is most likely not in an incomparability case at all. Rather than caring for his mother or leaving, he sees a third alternative—coming to Sartre for advice—which he seems able to compare to the other two and choose for its promise of helping with what will likely be an incomparability case when he goes back to it. In existential choice in general, a good first step is to attempt to compare. But in the incomparability case itself, the situation is different. One is in a genuine incomparability case only as long as one is unable to reach a comparative evaluation. As deep as the dilemma is, it is in one way like a game show. The outcome matters, but in a genuine incomparability case—with no Sartre to go to for advice—there is no alternative but to treat the situation as a Buridan case. Action from within an incomparability case must proceed in the same arbitrary way as in a Buridan case.

This means that, if the young man has to choose, he will have to make an arbitrary selection. That is, of course, part of Sartre's point in introducing such cases. Existential choices illustrate how much can rest on choices that we make without a reason. As we have already seen, the guise of the good can explain how we form an intention in such cases. The young man can justifiably decide to "simply pick," or let non-intentional processes make the selection, and then to follow whichever path he sets out on. There are no cases in which this strategy is unavailable; we are never forced to form an intention without forming a normative belief.

Though incomparability cases are not Buridan cases, it is not surprising that they can be accounted for in the same way. After all, as far as the Identity View is concerned, they raise the same problem. Both are troubling because it seems the person has no way to

---

already, in the first sense, picked. On the view I described in the previous section, the outward 'picking' may or may not be intentional; and the mental 'picking' can itself be disambiguated into a non-intentional mental event or act and the formation of an intention to follow through on the result. Picking a number is a somewhat more complex process than it can at first seem to be. We typically first form an intention to pick a number. Then we let our mind settle on one number—in, as before, a non-intentional way that we do not ourselves fully understand, though we do understand that it is the only way to do this. Then we intentionally keep the number in mind, and (usually) say it when prompted.

form a normative belief. In both kinds of case, when all attempts at deliberation fail, one can go on intentionally by to allowing oneself to go on non-intentionally and then going on intentionally. Just as on a game show, it does not matter that there are enormous differences between the alternatives, as long as those differences are inaccessible to normative belief. In typical incomparability cases, what allows us to proceed is our capacity to let ourselves act non-intentionally, and to intentionally follow through. As in Buridan cases, this is often what we intend under the guise of the good.

Let me summarize what I have argued. The problem faced by someone in a Buridan case can be put this way: "These are equally good. What do I do?" The problem faced by someone in an incomparability case can be put this way: "I can't decide. What do I do?" These are genuine practical problems, by which even a human being can, at least temporarily, find herself stumped or paralyzed. Such cases also present a theoretical problem. The second, theoretical problem can be put this way: "We form intentions and act intentionally in such cases. How do we do it?" This is a call for explanation that is especially pressing for particular theories, on which we can only intend or act intentionally when see one alternative as more worth pursuing than the others. The existence of Buridan and incomparability cases is a problem for acting beings; the resolution of these cases is a problem for theoretical views like the Identity View.

The solution I offered to the practical problem is: just pick one. That is, do what you do when you pick a number, or a card, or a cookie. Let yourself go on non-intentionally until one stands out, and go with that one. This is not the only possible solution, but I think it is a good one. It does not take much time or effort; it appeals to capacities we already have; and it can be unobvious, and helpful to someone who is gripped by the problem.

The solution to the theoretical problem is: we usually just pick one. That is, we let ourselves go on non-intentionally until one alternative stands out, and we intentionally pursue that one. We do this for good reason. It is better to pick than to remain stuck: better to take one can, or card, or cookie, than none at all. And there is no theoretically damaging tie here. What we intend to do, and what we do intentionally, is what we believe we ought to do: pick a date, and eat it.

Chapter 6: Evaluation without Motivation

We often find ourselves unmotivated. We believe we should get out of bed, but we lay around for minutes or even hours without so much as sitting up. We believe we should make a phone call or write an email, but, as we sometimes say, we "cannot get ourselves" to. Knowing how easily we can save another life, we believe we should send money to Oxfam or some other charity—but, we sometimes seem forced to admit, we have no intention to do it. We can fail to do, not only what others expect of us, but what we ourselves believe we should do.

These situations raise a cluster of practical problems. If we find ourselves unable to get out of bed in the morning, we might ask ourselves: what can we do to resolve this situation? Is there some thought experiment, or imaginative or associative technique, or piece of reasoning we can use to make it easier to stand up and start the day? We can also ask: how can we avoid the experience of such a situation? How can we avoid the unpleasantness, the sense of wasted time, of feeling frustrated, stuck, paralyzed, or guilty, that we often experience when unmotivated? And we might also ask: how can we avoid the consequences of being unmotivated? How can we avoid the actual waste of time, the damage to our work or our relationship with family and friends, or the actual loss of a young child's life that result from our failure to act? Both while we are not motivated and while we are, we can ask these practical questions. And of course, we can ask more general theoretical questions too. We can wonder what limits our lack of motivation places on the connection between evaluation and motivation—or more generally, what our lack of motivation shows about us as people.

There is also a more particular theoretical problem, directly related to the guise of the good. On what I have been calling the Identity View, an intention to do something is a belief that one should do it. When we lack motivation, we can at least seem to have the belief without the intention. An intention to stand up, and a belief that I ought to stand up, cannot be the same state if I can have the belief without having the intention. As a theoretical counterexample, lack of motivation complements *akrasia*, in which intention and normative belief conflict, and Buridan and incomparability cases, in which normative belief seems lacking. This time, the intention seems lacking. A defense of the Identity View must address all three kinds of counterexample. In this chapter, I offer an account of the third and last one: apparent cases of normative belief without intention.

The main task of this chapter is to show that lack of motivation is not a damaging counterexample to the Identity View. But as in earlier chapters, the larger goal is not purely defensive. It is to show that a guise-of-the-good view can account for the varieties of intentional activity, in a way that improves our understanding of each of them. So as I go along, I will try to show how the defense of the Identity View addresses the larger theoretical and practical questions about lack of motivation. If I am right, answering the

theoretical challenge will improve our conception of ourselves, and will help address the practical challenges as well.

As usual, it will help to try to consider the most difficult cases. In §I, I introduce several related kinds of case—depression, psychopathy, amoralism, and more ordinary failures—and say why I think the ordinary failures I already mentioned are the clearest and most difficult counterexamples. In §II, I consider three potential responses to the examples: denying the phenomenon, assimilation to *akrasia*, and appeal to conditional normative beliefs. I will argue that none of these responses address the full extent of the problem. In §§III-V, I offer a conception of lack of motivation centered on an analogy with fatigue. Drawing on empirical studies of willpower, I will argue that we can understand our failure to get out of bed as a failure in which we do intend to get up, but are hampered by a kind of exhaustion of our capacity to execute our intentions. On this view, central cases of lack of motivation are best understood as failures to do what one intends to do, rather than as failures to intend to do what one believes one should do. §VI draws out several further theoretical and practical implications.

I. Kinds of Example

Though "lack of motivation" captures roughly the range of cases I have in mind, "motivation" is a potentially misleading term. "Lack of motivation" can refer to a lack of desire, or to a failure to act, or to desires or intentions with relatively low "motivational strength." For our purposes, a deliberately broad range of examples could include all *failures to act on a normative belief*. We can take such inaction as the larger range of cases which this chapter attempts to address.[105]

The cases that mainly interest us here are those in which we seem to lack the *intention* to act in the ways we believe we ought to act. These are the cases that present apparent counterexamples to the Identity View. But even within this category, there is a great deal of variety. In this section I introduce several kinds of case, and raise some initial problems about their status as counterexamples.

[105] It might seem more natural to think of 'lack of motivation' as a failure to act on one's *intention*, rather than on one's belief. This would be in line with the view I go on to develop. But lack of motivation to act on a normative belief is the relevant problem case, and so the one I focus on here. On the other hand, there might also be cases of lack of motivation to refrain or to make an omission. If I like to count to 100, but believe, for whatever reason, that I should omit the number 25 today, I might still lack the motivation and not bother, allowing myself to say all the numbers by force of habit. Strictly speaking, then, some relevant cases of lack of motivation might not be best understood as failures to act. I leave out this complication in the text, though I hope that what I say can be applied to all failures to comply with our normative beliefs, whether by action or by omission.

1. Depression

Depression has been seen as a paradigm case, or even *the* paradigm case, of lack of motivation. We can be too depressed to get out of bed, or to make a phone call, or to work to help others. And it can be natural to think of depression as a clear case of an impairment that is specifically motivational rather than cognitive—an impairment in intention and action rather than in one's beliefs about what one should do. As Dancy (1993, 5) puts it: "The depressive is not deprived of the relevant beliefs by his depression; they just leave him indifferent." Roberts (2001, 43) writes: "Depressives, by their own lights, are not doing what *they* think they should. Depression, it appears, leaves one's evaluative outlook intact." So when a depressed person fails to act, the failure can seem to be distinctively practical, separable both in principle and in practice from a failure to believe one ought to act in some way. People with depression seem to, in Michael Stocker's phrase, "see all the good" in something but not be moved by it.[106]

But the complexities of depression make it a less than straightforward example. [107] People with depression do show significant cognitive differences from people who are not depressed. They have been found to describe what seem to others to be happy faces as neutral, for example, and neutral faces as sad. Some studies suggest that people with depression tend to attribute unfortunate events to stable causes, and rate their importance more highly than others do. It is then not surprising that a depressed person might tend to expect that desirable outcomes will not occur and unwanted ones will. Empirical confirmation on this point would not be a surprise to those of us who already associate depression with pessimism. And it might remind us of another, similar link. People with depression often have strikingly negative evaluations of their own self-worth, and of the world around them. They can find themselves thinking: "I don't deserve to live", or: "The world is a dump, anyway." Thoughts like these that can lead a depressed person to suicide, rather than to inaction. And they suggest, *pace* Roberts, that depression does not leave our

---

[106] The passage from Stocker (1979, 744) reads:

> Through spiritual or physical tiredness, through accidie, through weakness of body, through illness, through general apathy, through despair, through inability to concentrate, through a feeling of uselessness of futility, and so on, one may feel less and less motivated to seek what is good. One's lessened desire need not signal, much less be the product of, the fact that, or one's belief that, there is less good to be obtained or produced, as in the case of the universal Weltschmertz. Indeed, a frequent added defect of being in such 'depressions' is that one sees all the good to be won or saved and one lacks the will, interest, desire, or strength.

[107] For some representative scientific literature, see Beck (1963, 1987), Cook and Peterson (1986), Ellis (1987), McDermut et al (1987), and White et al (1992). For a useful survey and philosophical discussion of some of these empirical findings, see Bromwich (2008, 179-186). My brief discussions of depression and psychopathy are indebted to Bromwich's much more detailed treatment of these cases. A series of more recent helpful discussions of these and related phenomena can be found in Björnsson et al (2015).

evaluative outlook intact. Being "too depressed to get out of bed" can often be a case of being too unsure that life is worth living, or that anything good will come when the day starts. Those of us who are depressed can have a great deal of trouble "seeing all the good," both in ourselves and in the world around us.

There is also a methodological difficulty in considering depression. Talk of depression can be talk of clinical depression, as defined in psychology and psychiatry.[108] Or it can be talk of depression in a layperson's sense. Someone with no knowledge about clinical depression can say: "I'm depressed today", and 'depression' has in general become a household term rather than a purely technical one. Taking depression as a class of counterexamples calls for specifying what sort of depression is at issue. In either case, the specification faces a general association with changes in evaluation. In the clinical case, a consideration of depression must face empirical findings like the ones I have mentioned, and either pursue or await further relevant findings. In the ordinary case, it must face the everyday association with pessimism and low self-esteem.

One can try to avoid these issues by stipulating that the relevant kind or sense of 'depression' is precisely the one in which one's normative beliefs are intact but do not result in an intention. But, of course, even if the stipulation is granted, we have not come any farther in specifying the relevant kind of counterexample. We would then need to start again in the search for a compelling case. And we would not avoid the question of whether "depression" in the relevant sense exists.[109]

A simple mention of depression as a counterexample leaves a great deal of room for further specification and alternative interpretations of the case. So although it is worth keeping depression in mind, and although many relevant examples can be examples of depression, it is also worth asking whether there are other, clearer and harder, counterexamples.

2. Psychopathy

Like people suffering from depression, psychopaths have been thought to be paradigmatic examples of a lack of motivation to do what they believe they should. They lie, manipulate, and even kill, apparently with an intact moral understanding. As Doris and Stich (2005, 124) put it: "Psychopaths…appear to *know* the difference between right and wrong but quite generally lack motivation to do what is right."

Psychopathy too is a delicate example. As a folk intuition, such a view about psychopaths carries little weight. Even the thought that psychopaths *appear* to be this way

[108] Appeals to clinical depression in rejections of guise-of-the-good views are rare—but see Mele (1996). For further discussion, see Bromwich (2008, 172ff).

[109] Though the passage from Stocker (see n.1 above) does not explicitly use 'depression' in a stipulated sense, it can illustrate how unhelpful that would be. If we were to call the phenomena Stocker mentions "depressions," talk of depression would no longer be introducing an independently characterized or useful kind of example. Even in Stocker's description, it is the other characterizations that introduce the relevant descriptions, with the label "depressions" added afterwards.

is a piece of hearsay, or an uneducated guess, about a particular pathology. And it is controversial at best. As with depression, the clinical findings are more complex, and point in part to cognitive impairments. Psychopaths have been found to have difficulty with affective language, to such an extent that some psychologists conclude that "psychopaths…often have difficulty in understanding and using words that for normal people refer to ordinary emotional events and feelings" (Intrator et al 1997, 101).[110] At the same time, psychopaths have been found to have an unusual degree of difficulty with abstract word processing tasks (Kiehl et al 1999), suggesting that they have cognitive impairments related to both affective and abstract concepts. That actual psychopaths do have unimpaired normative beliefs is thus controversial at the very least. As Nichols (2002, 293) puts it: "Recent evidence suggests that psychopaths really do have a defective understanding of moral violations." Although the possibility remains open, it is not easy to find a compelling and empirically accurate example in which a psychopath clearly does have a normative belief that nevertheless fails to motivate her.

3. Amoralism

Apart from clinical cases, there is also the ordinary amoralist: someone who understands what morality requires, but simply does not care. Many of us have met someone who insists she has no interest in living a morally good life—someone who, as Brink (1989, 46) puts it, "recognizes the existence of moral considerations and remains unmoved." Even without a psychiatric diagnosis, such a person can seem to be a classic case of absence of intention to do what one believes one ought to.

But again, as with depression and psychopathy, there are several complications. First, we should ask whether such a person believes she *ought* to do what morality requires. If she thinks of moral requirements as mere social conventions, or as a set of commands whose authority we can question, she might not have a normative belief in favor of doing what morality asks of us. She might in fact believe we should do something else: follow our own pleasure, for example, or make our own rules, or 'live in the moment'. And she might intend to do just that. In that case, she would intend to do exactly what she believes she should.

Second, we should ask whether to believe such a person. A self-proclaimed amoralist might not be an actual amoralist. An amoralistic rant can be a piece of rebellious posturing, made by someone who an hour later is full of moral indignation at a driver who cuts her off on the highway. The deeper question is whether there are any amoralists at all—whether it is possible to not care at all about what one ought to do. Taking amoralism as a counterexample would call for a defense of this possibility.

Third, even granting the possibility of total indifference to what one ought to do, we should ask whether someone so indifferent could have normative beliefs to begin with. Just as we can doubt that psychopaths understand normative terms well enough for these

[110] For discussion of this and other, convergent findings, see Bromwich (2008, 208ff).

terms to play a role in their beliefs, we can doubt that an amoralist could have the normative beliefs to which she would (if she had them) be indifferent. These beliefs would have at best a very limited role to play in her psychology, and might not be recognizable as beliefs at all. This is another difficult problem about this kind of example, which requires both the presence of normative belief and a lack of motivation by it. Far from an immediately compelling counterexample, amoralism can easily lead to, as Svavarsdóttir (2006, 165) puts it, "a stalemate of conflicting intuitions."

Fourth, we can distinguish global and local amoralism. A global amoralist would not care about anything she believes she (morally) ought to do. Or rather, she might happen to care—she might, by coincidence, both want to eat dinner and believe she should. But it would not matter to her that she should, in any context. Her normative beliefs in general would have no significance for her. A local amoralist would be left cold by *some* of her moral or normative beliefs, while caring about others. Either kind of amoralism has its difficulties. Genuine global amoralism is difficult to imagine, and raises doubts about the presence of normative belief. Local amoralism, on the other hand, adds little to a more general characterization of lack of motivation. We began with lack of motivation to do what one believes one ought. If local amoralism is itself a lack of motivation, in some particular cases, to do what one believes one ought, it is simply another name for lack of motivation. It is not yet a particular kind of example.

4. Ordinary Failures

None of the three kinds of example considered so far in this section include the ordinary failures with which this chapter began. Ordinary, happy, non-depressed, non-psychopathic people regularly fail to get out of bed, or make a phone call, or donate money even when they believe they should. And I think these are the best examples to consider. Their existence is difficult to deny; they are recognizable and widespread; they require no global failure on the part of the person undergoing them; their significance cannot be called into question by scientific controversies about clinical conditions; and they offer a direct and intuitively troubling challenge to the Identity View. One cannot claim that they depend on ignorance about a particular pathology. And, as we will see, it is not easy to deny that they involve genuine normative belief. These examples are thus, on the whole, both clearer and harder. They are vivid and recognizable, and difficult to avoid or explain away.

This is not to say that the other cases exclude these. There is no reason to deny that people who are depressed, psychopathic, or (professed) amoralists are also subject to lack of motivation in the more ordinary ways. And depression, psychopathy, amoralism, and other particular conditions might include distinctive variants on the ordinary cases. A theoretical response to lack of motivation must be general enough to address any potential counterexample. This is why I include all four kinds of example as at least potentially problematic, and why I do not claim that the list is exhaustive. I include them, both to describe a range of examples, and to explain why—despite the natural and widespread appeals to the others—I will focus on the ordinary cases. To the extent that the other

examples are examples of failing to do what one believes one should, a general account of ordinary lack of motivation should apply to them as well.

II. Initial responses

With a range of examples in mind, we can now ask: how can the Identity View account for lack of motivation?

1. Denying the Phenomenon

One response to the cases is to deny that we ever fail to act on our normative beliefs. If we really believe we should get out of bed, this response would say, we do it. And if we do not do it, we do not really believe we should.

Any defense of the Identity View must deny that we can lack an intention to do what we believe we should do. On the Identity View, the belief is itself the intention. What we are considering here is a further denial. It is an attempt to avoid accounting for a kind of inaction that is problematic for the Identity View, by denying the possibility of the inaction itself.

As with Buridan cases (see Chapter 5), such a denial is difficult to maintain. It seems to fly in the face of the apparent fact that we do fail to act on our normative beliefs. We do seem to do nothing even when we think we should get up, or make a phone call, or send money to charity. It is hard to see why we should think this does not happen.

Part of the challenge is to explain what else, other than a normative belief, the inactive person would have. What is the person saying or thinking when she says or thinks: "I should get out of bed"? According to the denial, if she does not get up, she does not really believe she should. So what does she believe?

One possibility is that she is not using "should" in a genuinely normative sense. She might think "I should get out of bed" as she would think: "One should eat salad with the smaller fork." That is, she might be using the "should" of social convention. In that case there would not be a damaging problem. We can believe that one should eat salad with the smaller fork, but not do it, because we believe that, in the normative sense, it is not true that we should follow social convention in this case.[111]

Another possibility would allow that she can use "should" in a normative sense. If two teenage siblings believe their parents are being unreasonable, one can tell the other:

[111] It might be objected that the 'should' of social convention has its own kind of normativity, making it incorrect or at least misleading to deny that it is genuinely normative. If we grant this, the point can be put differently. To believe that one should eat salad with the smaller fork, with the 'should' of social convention, is not yet to believe that, all things considered, that is what one ought to do. That the 'should' of social convention is decisive for action is a further thought that one may or may not have in a particular case. One can believe 'I should get out of bed' in this social convention-like way, without believing one ought to follow the convention.

"We should be home by nine," as a way of reporting what her parents said. In that case "We should be home by nine" is a kind of indirect discourse, which does not report the sibling's own beliefs. "I should get out of bed" can be said this way as well. According to the norms of our culture, one might implicitly say, I should get out of bed. But one can add: I do not accept the norms of our culture. And so, one can say: "I should get out of bed. But I'm tired and I really like it here, and it's really okay if I stay for a while."[112]

Either of these two possibilities can be expressed by describing a person as saying or thinking: "I 'should' get out of bed." The single 'scare' quotes, or inverted commas, single out "should" as playing less than its usual normative role. And the possibility of these inverted commas uses offers a way to explain the appearance of normative belief. What seems to be normative belief can in fact be inverted commas belief—a belief that one 'should' do something, where the use of "should" is not genuinely normative. We can call this *the inverted commas reply*.[113]

Nevertheless, a recognizable alternative to normative belief is not enough. The denial must deny that we can *ever* believe we should get out of bed without actually doing it. Emphasizing the possibility of one interesting phenomenon—inverted commas belief—does not rule out the possibility of another: genuine normative belief in such cases. Here there are at least two further problems. First, there is what, in Chapter 4, I called the error attribution problem. Someone who fails to get out of bed can insist that she does believe she should, in as robust a sense of "should" as one can imagine. A theory that denies the possibility of failing to act on a normative belief must say that she does not believe it. Why should we believe the theory, and not her? Faced with a person's insistence that she does have a belief, it is hard to see what the argument for the denial would be. In the case of akratic action, I argued that the Identity View attributes error only in *denying* that one has a normative belief; and that it can point to a conflicting belief to account for the error. Here the problem is worse on both counts. The error is an error in attribution rather than in denial, since the person does believe she has a particular normative belief. She can insist: "But I do believe I should get out of bed. I'm sure of it!" And it is unclear what would account for her error. If she does not believe she should get out of bed, why does she believe she believes it?

Second, the denial seems to deny too much. More specifically, it denies even the possibility of *akrasia*, at least in some cases. According to the denial, if we really believe

---

[112] One way to understand the "should" of social convention is as itself a kind of indirect discourse. On such a view, when we say that we should eat salad with the smaller fork, we do not use "should" in a different sense. Instead, we mean either that we really should—in which case we endorse the convention—or that, according to convention, we should. I do not mean to take any position about the "should" of social convention here. The point is only that it offers a way to explain the appearance of genuinely normative belief in cases of inaction.

[113] For a classic discussion of "inverted commas" statements, see Hare (1952, 124-6, 163-5). My use of "inverted commas" is broader than Hare's, who includes only the first of the two possibilities I mention, calling the second a "conventional use" rather than an inverted commas use (125). I include the second possibility under "inverted commas," both for simplicity and because the scare quotes or inverted commas are appropriate in both cases.

we should get out of bed, we do it. Akratically reading a book in bed is then ruled out as well. If we believe we should eat a healthy meal, on this view, we could not eat a dessert that we ourselves believe is unhealthy. We would be unable to do anything incompatible with something we believe we ought to do. This denial of a wide range of akratic action flies in the face of our experience. As we saw in Chapter 4, it also denies the possibility of conflicting normative beliefs, since one could not act on both of the conflicting beliefs. In other words, it both faces the error attribution problem, and violates what I called the Conflict Constraint on an understanding of akratic action. The denial is thus implausible, both in itself and in its theoretical consequences.

A denial that we can fail to act on our normative beliefs is an extreme position, and difficult to maintain. I think it is an extreme move motivated mainly by the demands of a theory. As I will argue in §III, it is an unnecessary move. Even on the Identity View, we can explain the possibility of failure to act, rather than denying it.

2. *Assimilation to* Akrasia

Denying the possibility of failing to act on normative belief tends to deny the possibility of akratic action. But then, one might think, this might be because the failure to act is itself a kind of *akrasia*. Could we not account for these failures the same way we account for the possibility of acting akratically?

This thought becomes more attractive when we consider intentions to refrain. We have the capacity not only to intend to eat dessert, but to intend *not* to eat it. We can think about whether to eat it, weigh the considerations for and against, and form an intention not to. We can similarly intend not to get out of bed, or not to make a phone call, or to refrain from donating to charity. We can have intentions whose content is to, in a certain respect, do nothing. So when we believe we should get out of bed, and fail to do it, this can be because we intend not to. Staying in bed can be an intentional refraining, prompted by a conflicting motivation. Laziness, discouragement, or fear can lead us to intend not to do anything, even when we believe we should act.

Just as some apparent expressions of normative belief can be the indirect discourse of inverted commas, some failures to act on normative belief can be akratic. We can and, I think, should allow that akratic refraining is possible and even widespread. But as with inverted commas belief, I do not think we should conclude that akratic refraining accounts for all the failures.

To see this, it helps to distinguish akratic and non-akratic failure to act on a normative belief. To akratically fail to do something, we can say, is to intentionally not do it, while believing one ought to do it. To fail non-akratically would be to believe one ought to do something, and to not do it, without intentionally not doing it.

To assimilate failure to act on a normative belief to *akrasia* is to deny the possibility of non-akratic failure to do what one believes one ought to do. It is to treat all of the failures as akratic. Although this is not a denial of the phenomenon of failing to act on

a normative belief, it is a denial of another, more particular phenomenon. It is a denial of the possibility of non-intentional failure to act on a normative belief.

Gosling (1990, 190) gives a helpful description of non-akratic failure. He writes:

> There is not a counter-purpose, as with fear, but a lack of interest in anything so energetic as decision or action. This makes it unnatural to describe the agent's failure as deliberate, or even intentional, although it will commonly be true that the agent is knowingly failing.

Gosling's phrase "lack of interest in anything so energetic" helps to capture why staying in bed is such a natural example. When one stays in bed, we might say, one is often uninterested in anything so energetic as brushing one's teeth, having breakfast, or going to work. And although we do sometimes intentionally conserve our energy by staying in bed, there is so far no reason to deny that we can also stay in bed without ever deciding to. We might not get around to, as Gosling puts it, anything so energetic as decision one way or the other.[114]

Assimilating all failure to act on a normative belief to *akrasia* denies the phenomenon of non-akratic failure. And it ignores a distinct and potentially more difficult problem. It is often difficult to explain a failure to act on a normative belief by a conflicting motivation, as we usually would in akratic action. We can want to know how it is that we can fail, even without the conflict. Leaving non-akratic failures unaccounted for then threatens to make the problem look too easy, by ignoring an independent source of doubt about the Identity View. It seems possible that what stops us from acting on a normative belief is simply a lack of motivation, rather than a conflicting motivation. To do justice to the difficulty of the problem, a defense of the Identity View should have something to say about this kind of case as well.

As with denial of the possibility of *any* failure to act on a normative belief, I think the denial of the possibility of non-akratic failure is poorly motivated and unnecessary. We do not need to deny *this* phenomenon, either. I will soon explain how we can account for it. But first, I want to consider one other line of response.

---

[114] I do not mean to suggest that intention requires decision or effort, or that everything non-intentional is also not energetic. I might intend to brush my teeth tonight, without any noticeable expenditure of energy in the intention. Conversely, I might believe that I should stop reading and go to bed, but continue reading compulsively or in some other way non-intentionally; or recognize that I should stop heckling a friend (it was funny at first, but now I've gone to far), but be unable to get myself to stop. These can be cases of lack of motivation as well, albeit more complex, and would count as what I called 'ordinary' failures, rather than depressive, psychopathic, or amoralistic ones. The phrase "lack of interest in anything so energetic" is meant to capture, not an essential lack of energy, but a failure to do as much as form a corresponding intention. On the other hand, I will describe in the next section a way to understand these failures that does essentially appeal to a lack of a kind of volitional energy. In that sense, compulsive reading or heckling would indeed be explainable by a lack of energy.

3. Conditioned Value

Sergio Tenenbaum (2007, Chapter 8) offers a conception of "accidie," or, roughly, lack of motivation, in defense of a guise-of-the-good—or, as he puts it, "scholastic"—view. He begins by insisting on the theoretical inadequacy of assimilating inaction to *akrasia* (2007, 283):

> Someone who suffers from *accidie* supposedly still accepts that various things are good or valuable but is not motivated to pursue any of them. This phenomenon seems harder to accommodate within the framework of the scholastic view than *akrasia* because here there is not a different (even if lesser) good that motivates the agent. At any rate, our way of explicating *akrasia* by means of the scholastic view does not seem to have any straightforward application to the cases of *accidie*; it is quite implausible to say that an agent who is in the state of *accidie* is somehow persuaded by an appearance of the good of, say, "staying put."

Tenenbaum focuses on a slightly different phenomenon from mine: judgments of goodness or value, rather than normative belief. But his aim is roughly similar. He believes that even with an account of *akrasia* in place, a separate treatment is called for in the case of failure to pursue what one sees as valuable. And he offers such a treatment, centered on a relation he calls "conditioning" (290):

> *Strong Conditionality*. *C* strongly conditions an evaluative perspective for an agent *A* if and only if, for every *O* conceived to be good from that perspective, *A* should judge *O* to be good only if *C* obtains.
>
> *Weak Conditionality*. *C* weakly conditions an evaluative perspective for an agent *A* if and only if, for some *O* conceived to be good from that perspective, *A* should judge *O* to be of lesser value if *C* does not obtain than if *C* obtains.

As Tenenbaum explains, relations of conditioning can be quite specific. Someone whose grandfather dies can lose interest in fishing, despite continuing to see fishing as valuable, because it is no longer the same without his grandfather there. In this case O is fishing, and C is the presence or participation of the grandfather. Even within the evaluative perspective from which the grandchild conceives of fishing as good, it might be that he should judge fishing to be of lesser value without his grandfather. He himself can believe in such conditioning. If he believes he should judge fishing to be of lesser value without his grandfather, he believes in a relation of weak conditionality in this case. In the extreme case, he might believe that without his grandfather, he should no longer judge anything to

be good. This would be a belief in strong conditionality.

We can believe in conditioning in a wide range of circumstances, both very particular and very vague. A depressed person, for example, can judge some or all of his life to no longer be good, or no longer be as valuable, "given that I feel this way", "given the kind of person I am," "given that my life has turned this way," or "given all that has happened around me"(293-4). We can take a wide variety of evaluative perspectives to be conditioned by a wide variety of states of affairs. We can believe in conditioning rightly or wrongly, vaguely or with precision, hesitantly or with conviction, and in both ordinary and philosophical contexts.

According to Tenenbaum, "the best way for a scholastic view to accommodate *accidie* is by means of this relation of conditionality. We can say that the agent in a state of *accidie* takes certain evaluative perspectives to be conditioned by certain states of affairs that do not obtain"(2007, 293). As with conditioning in general, the conditioning states of affairs can be conceived of in various and often very vague ways, such as "Given that I feel this way." Though the condition can vary, the idea is that the person takes *some* such conditioning relation to hold, and the condition to not be met. Though such a person "judges certain things to be valuable, he thinks that some of the facts we gave constitute a violation of a condition of his evaluative perspective and thus a violation of a condition of their being good or worth pursuing"(294).

Tenenbaum's description can be apt for many cases, including many cases of depression. But I think it is also a kind of denial of the phenomenon. Since the condition of something's being good or worth pursuing is seen as violated, the person sees that thing as *not* good or worth pursuing in his case. It seemed we were interested in someone's believing that something *is* good or worth pursuing—in my version, believing that she ought to do something—and still not acting in accordance with that belief. On Tenenbaum's view, it is essential that there be a change in evaluation. Even with "weak" conditionality, the person must see what she fails to pursue as having lesser value than it would in different circumstances.[115]

Tenenbaum's appeal to conditioning thus leaves in place the doubt with which this chapter began. It seems possible and even common for us to believe that something is valuable, good, and worth pursuing, believe that we ought to pursue it, and still not pursue it. To put it differently: for any conditions we can place on these evaluations, it seems possible, at least on the face of it, for us to see the conditions as met, and still not act. One

[115] Though I leave out this complication in the text, the details of Tenenbaum's characterization leave it to some extent open what sort of evaluative change is required. He writes: "We should see the agent who suffers from *accidie* as being *committed* to a certain relation of conditionality. The proposal does not require that the agent be able to immediately describe or even assent to the attitude ascribed to her"(295). One way to understand this commitment is as a kind of introspectively inaccessible belief, as in the psychoanalytic cases I discuss in Chapter 2. But the "commitment" may not be a belief for Tenenbaum, who often resists characterizations of evaluative states as beliefs. See his (2007), Chapter 2, discussed above in my Chapter 1. As I go on to argue in this chapter, we do not need any explanation in terms of conditionality to make the agent intelligible.

can believe that given the kind of person one is, and given how much work there is to do, and so on, one really should get out of bed. And still one stays in bed. To put it yet another way it seems that there are some cases in which no conditioning relation of the kind Tenenbaum describes has its condition unmet, and yet one is still in a state of *accidie*. We do not yet see either how this is possible, or why it would not be possible.

There is another difficulty with explanations in terms of conditioning relations. Such explanations are, broadly speaking, cognitive. They explain failure to act by appeal to a person's conception of the way a state of affairs sets a condition on an object's being valuable or worth pursuing. But if that is right, it becomes difficult to distinguish *accidie* from ordinary decision-making in general. When we make a decision, we normally consider alternatives that seem to us to have at least some value. Courses of action that have nothing at all to be said for them rarely, if ever, make it into the initial range of options to consider. And then we decide against one or more alternatives, because their being worth pursuing depends on some state of affairs that does not obtain. We decide against taking a walk, because it is too cold; or turn down a job, because the pay and benefits are not as high as we had hoped; or stay in bed, because we slept less than usual and do not need to rush this morning. If that is *accidie*, then it seems that most or all actions are examples of *accidie*. But then we lose our grip on the distinctive phenomenon to be explained.[116]

Judgments of conditioned value can seem theoretically promising if one sees a need to locate the explanation for *accidie* in a person's evaluative outlook. Tenenbaum writes that on his view, unlike views that reject the guise of the good, "we need not see *accidie* as the result of a surd lack of 'oomph' on the part of our evaluations, as the result of something completely external to how the agent views the world"(294). But I think such a motivation is misguided. A guise-of-the-good view must reject the idea that action, or intention, or desire can be the result of something completely external to how a person views the world. But a non-intentional failure need not result from something in a person's evaluative outlook. We can intend to get to work on time, because we believe we should, but fail because we get stuck in traffic. Or we can intend to get to work on time, because

---

[116] This difficulty comes closest to the surface in Tenenbaum's spelling out of his distinction between full-blown, hesistant, and inconsistent accidie (296):

> An agent who engaged in vicious behavior in the past but now, on account of accepting some kind of Kantian view of the relation between virtue and happiness, does not find her happiness worth pursuing, would be suffering, on this account, from full-blown accidie….She would be capable of seeing that if certain desirable conditions were to obtain, her happiness would be worth pursuing, but given that these conditions do not obtain, her happiness cannot be judged to be good.

Tenenbaum's example can be seen as someone who straightforwardly believes that some end—in this case, her own happiness—is not worth pursuing. But, of course, we all believe various things are not worth pursuing, because of some state of affairs that does not obtain. If that is full-blown *accidie*, we are all suffering from it all of the time.

we believe we should, but fail because we cannot muster the motivation to get out of bed. Though the second failure must in a sense be attributed to 'internal' causes, it might not be attributable to a person's evaluation. Even on a guise-of-the-good view, there is no need for a *nonintentional* failure to itself be 'under the guise of the good'.

In the rest of this chapter, I will offer an account of failure to act on one's normative beliefs that is consistent with the Identity View, requires no failure of evaluation, and sheds light on the psychology of motivation. But rather than rely on intuitions about motivation, I turn first to some empirical work in psychology.

III. Baumeister's Strength Model

To illustrate the connection to fatigue, consider a series of influential studies by Roy Baumeister and his colleagues.[117] Each study gave experimental subjects two consecutive tasks, and measured the effects of engaging in the first task on performance in the second one. People asked to resist tempting chocolates, for example, show impaired performance on a subsequent puzzle task. Effortful regulation of one's mood—in either a 'positive' or a 'negative' direction—decreases subsequent time spent on a handgrip squeezing task. Suppressing thoughts of white bears decreases time spent solving anagrams. In general, experimental subjects are less willing to put extended effort into a task when they have just completed a different, even if seemingly unrelated, effortful activity. Baumeister calls this effect "ego depletion… a temporary reduction in…capacity or willingness to engage in volitional action…caused by prior exercise of volition"(Baumeister, Bratslavsky, Muraven, and Tice 1998, 1253).

Baumeister and his colleagues—and, increasingly, psychologists in other laboratories—have found ego depletion across a wide variety of activities. Outside the laboratory, coding of autobiographical stories shows an association between self-regulation failure and prior self-regulation. In consumer behavior, extended shopping has been found to reduce willingness to compromise and make decisions. Overall, Muraven, Tice, and Baumeister (1998, 786) conclude, there is "converging evidence from several very different research methods…: After people exercise self-regulation, they are subsequently less capable of regulating themselves, at least for a short time." An independent meta-analysis of 83 studies in a range of laboratories found "a significant…ego-depletion effect…generalizable across spheres of self-control"(Hagger, Wood, and Stiff 2010, 515).

[117] See Baumeister (2002, 2003, and 2012); Baumeister, Bratslavsky, Muraven, and Tice (1998); Baumeister and Exline (1999); Baumeister and Heatherton (1996); Baumeister, Sparks, Stillman, and Vohs (2008); Baumeister and Vohs (2007); Baumeister, Vohs, and Tice (2007); Muraven and Baumeister (2000); Muraven, Tice, and Baumeister (1998); Schmeichel, Vohs, and Baumeister (2003); and Vohs and Baumeister (2004). Much of the research is summarized in Baumeister and Tierney (2011). Since I will quote from a range of sources with varied co-authors, I will, for ease of presentation, continue in the text to treat the central ideas as Baumeister's. But it is worth remembering that his project is very much a team effort.

For Baumeister, the topic of these studies is self-regulation, or the "capacity to alter or override one's responses, including thoughts, emotions, and actions"(Baumeister 2002, 129). We regulate ourselves when we, for example, suppress thoughts of white bears, or persist in squeezing a handgrip. Baumeister finds his studies to support the introduction of a notion of strength into a conception of self-regulation. As Muraven, Tice, and Baumeister (1998, 775) describe it,

> A strength model of self-regulation depends on three points. First, the process of self-regulation consumes some resource, leaving it depleted afterward. Second, success at self-regulation depends on the availability of this resource, and possibly self-regulation may be a linear function of this resource. Third, all forms of self-regulation require some such resource, and indeed they may all draw on the same resource. These assumptions furnish the relevant prediction that an act of self-regulation will be followed by poorer self-regulation even in other, quite different, spheres.

These three points can be seen as different aspects of the idea that a kind of depletion impairs performance on the second task. The first point introduces the notion of depletion. A self-regulation task leaves us with less volitional resources, or less strength for self-regulation. The second point—that success at self-regulation depends on the availability of this resource—takes depletion to be explanatory. Success, and therefore failure, on a subsequent task depends on and can therefore be explained by the availability of the resource. Depletion can account for failure. The third point makes a claim of generality. *All* forms of self-regulation require some such resource, and perhaps even a single resource. Self-regulation in general, Baumeister thinks, requires and consumes a kind of strength. In other words: (1) Self-regulation depletes a resource, and (2) depletion affects self-regulation (3) in all its forms.[118]

In part to emphasize the notion of strength, Baumeister often draws an analogy between the resource in question and the strength of a muscle. The analogy is highlighted in some of his titles: e.g., "Virtue, Personality, and Social Relations: Self-Control as the

[118] The three-point strength model also suggests distinct but closely related aspects of the idea that it is the *ego* that is depleted. Baumeister, Bratslavsky, Muraven, and Tice (1998, 1253) note that "the notion that volition depends on the self's expenditure of some limited resource was anticipated by Freud,… [who] thought the ego needed to have some form of energy to accomplish its tasks and to resist the energetic promptings of id and superego." For Baumeister, talk of ego depletion is in part an homage to a Freudian energy model. The ego uses energy, a resource that can be depleted. When it is depleted, it is less able to accomplish its tasks. Moreover, as Freud knew, "ego" is the Latin for "I". If self-regulation consumes a person's volitional strength or energy, the notion of an ego goes naturally both with the focus on being regulated by oneself, and with the notion, suggested in Baumeister's third point, of a single resource used in a wide variety of activities. At the same time, the use of "ego" suggests a simpler thought: that after a difficult self-regulation task, *I* am depleted. I myself have less energy, and am less able to accomplish my tasks. Though the word 'ego' does not itself explain or justify that thought, it offers a way of expressing it.

Moral Muscle"(Baumeister and Exline 1999); "Self-Regulation and Depletion of Limited Resources: Does Self-Control Resemble a Muscle?"(Muraven and Baumeister 2000); and "Self-Control: The Moral Muscle"(Baumeister 2012). As Baumeister notes, muscular and volitional performance share three central features: temporary depletion, long-term improvement with practice, and adaptability for various tasks (Baumeister and Exline 1999; Muraven and Baumeister 2000; Baumeister 2012). Like our muscles, he might say, our wills can become depleted, strengthen over time, and be used for and exhausted by various, seemingly unrelated tasks.[119]

One advantage of this muscle-like strength model is that it is already familiar to us in ordinary contexts. According to Muraven, Tice, and Baumeister (1998, 774), "The strength model of self-regulation is implicit in the traditional concept of *willpower*." It is not difficult to say why: we tend to think of willpower as a kind of resource or 'power' that can be depleted, on which our success often depends, and on which self-regulation in general will draw. Of course, a connection to the notion of willpower can seem to be a disadvantage if that notion is thought to be especially obscure or problematic. Baumeister acknowledges that, in general, "Energy models are far out of fashion in modern psychological theory"(2002, 132). Nevertheless, he thinks, empirical support for an energy or strength model shows that "The folk notion of willpower is not far off the mark"(2012, 113). He describes himself as "bring[ing] back the Victorian notion of willpower…. Willpower may have an unappealing, Victorian reputation. But it is simply a matter of using one's physical and mental energy to reach one's goals and get the most out of

[119] Though these are the three Baumeister emphasizes, one might, more speculatively, suggest three further features of the muscle analogy. Fourth, there is the potential for atrophy. We know that muscles can atrophy when left unused; whether volitional capacities do as well is an area for further investigation. Fifth, behavioral evidence takes priority over self-report evidence in both cases. Though Baumeister does not explicitly include the priority of behavioral evidence in making the muscle analogy, he does suggest it in the case of ego depletion. Baumeister and Exline (1999, 1181) write: "It is unclear how valuable self-report measures can be. Indeed, once social desirability biases are corrected, there may be little or nothing useful or valid in self-reports about ego strength, implying the need for behavioral measures." Sixth, as Baumeister eventually emphasizes and as I will go on to discuss, deliberate conservation, as opposed to total exhaustion, plays an important role in diminished performance in both muscular and volitional tasks.

Relatedly, Baumeister takes a marked interest in the physical underpinnings of ego depletion, and particularly in its correlation with decreased levels of blood glucose. The connection to blood glucose has itself become the object of an increasing body of research, some of whose findings are striking. The need for blood glucose for self-regulatory success makes dieting a uniquely difficult exercise of willpower, creating what Baumeister calls "the perfect storm of dieting"; see Baumeister and Tierney 2011, Chapter 10. And a lack of food can have a striking effect on a wide range of volitional activity. Baumeister 2012 reports a study of parole hearings that found the chance of parole in a case heard just after lunch to be near 65%, and the chance of parole just before lunch to be near 0%. Baumeister defends a default of refusal to parole that some may find questionable; but the effects of blood glucose are nevertheless a promising area for further study.

life"(2012, 112-115).[120] For Baumeister, the extent to which self-regulation tasks impair performance on subsequent, apparently unrelated tasks both illustrates and supports the thought that these tasks draw on a kind of energy.

Baumeister's studies, I will argue, offer the seeds of a compelling explanation of many cases of failure to act on one's normative beliefs, without requiring a failure to have the relevant intention. But showing this is not easy. To begin with, the studies themselves can be the object of several distinct misgivings. In some cases, I believe that Baumeister himself offers a compelling reply; in others, one can be given on his behalf; and in others, the objection calls for a revision of his strength model. I begin with what I take to be the most easily answerable objections, and then move on to more difficult ones.

(1) Triviality

Do Baumeister's ego depletion studies show anything significant? They can seem trivial: a paradigm case of psychology proving the obvious. Of course a tiring self-regulation task makes it more difficult to perform another tiring self-regulation task. This is hardly the discovery of the century. On the contrary, it would be surprising if one such task had no effect on the next one. That it does have an effect, one might think, is too obvious to be interesting. We can call this *the triviality objection*.

The triviality objection is testable. In Muraven, Tice, and Baumeister (1998), Study 2, the initial depleting task asked subjects to try not to think about a white bear while writing down their thoughts on paper. A second task measured the duration of persistence in attempting to solve an anagram that, unbeknownst to the subjects, was in fact unsolvable. Typically for Baumeister's studies, neither task was especially exhausting, and the second task was chosen to be different enough from the first that it "would have no apparent relation to the initial manipulation"(779). The experimenter questioned each participant at the end of the study, and found that "No people believed that their performance on the first part of the study had any impact on their performance on the second part of the study"(780). For Baumeister and his colleagues, this was a way participants were "probed…for suspicion regarding the experimental manipulations"(780). But at the same time, although the results do not call attention to this, the questioning was also a way to test the extent to which the results of the experiment would be surprising. *All* of the participants reported that they did not believe their performance on the first task affected their subsequent performance.[121] With a pair of tasks that do not seem

[120] I take Baumeister as the most representative, influential, and sophisticated recent proponent of a strength-, energy-, or willpower-centered theory of self-regulation, though, as he insists, he is not the first. Mischel (1996) and others have also "proposed that the traditional notion of willpower needs to be revived to account for delay of gratification and similar patterns of self-regulation"(Baumeister et al 1998a, 774).

[121] The study had 58 participants, and dropped 7 who were not native speakers of English from its data analysis. The 51 remaining subjects were randomly assigned to one of three conditions—thought suppression, a control group with no instruction, and instruction to think about white bears as much as they could—with "17 in each condition"(780, note to Table 2). In other words, 17 out

significantly related or especially difficult, this expectation is itself unsurprising. But it does support the view that Baumeister's results are surprising. Indeed, they would have surprised every single participant. None of the participants expected an ego depletion effect across such different and comparatively easy tasks. It is striking that, as Baumeister might put it, one moderate use of willpower leaves less willpower even for a very different activity.

(2) Philosophical Triviality

Although the experimental results can surprise us about the extent of ego depletion effects, one can doubt that they have any significant philosophical implications. We might learn something about the extent or frequency of motivational effects of one task on another. But, one might think, we cannot learn anything philosophically important from them about our capacity for self-regulation, our failure to act on our normative beliefs, or anything else of philosophical interest. We can call this *the philosophical triviality objection*.

This objection, too, underestimates the power of empirical inquiry. Philosophical views about the capacity for self-regulation can themselves be testable. If our philosophical theories have empirical implications, empirical findings will, by the same token, bear on those theories. Muraven, Tice, and Baumeister (1998, 775-6) note some of these empirical implications, especially in predicting the effect of performance in an initial task on performance on the second one. We can ask: is the capacity for self-regulation a kind of knowledge, or a skill, or a limited but constant capacity, or a kind of resource? If the capacity for self-regulation is a kind of knowledge, its exercise seems likely to prime or activate it, acting as a reminder. The reminder, it is natural to think, will tend to improve performance on an immediately subsequent activity. If the capacity for self-regulation is a skill, its exercise might not have any immediate effect on the next exercise. If it is a limited but constant capacity, exercise might affect simultaneous activities, but not subsequent ones. But if it is a resource subject to depletion, it should show at least short-term decrease after exercise, with lower success and persistence rates on the second task. These predictions can be summarized in the following table:

The capacity for self-regulation	**Predicted effect of 'depleting' task on performance in second task**
Knowledge	Improvement
Skill	No effect
Constant capacity	Effect only on simultaneous tasks
Depletable resource	Impairment

of 17 people in the thought suppression condition would be surprised to find that their suppression of thoughts about white bears affected their performance on the anagram task.

For each conception in the first column, one can in principle develop an alternative view of the prediction. There can, at least in principle, be more sophisticated versions of each of these theories that might predict different ouctomes. But the basic point still holds: these views of the nature of the capacity for self-regulation are not inert with respect to empirical prediction. Whether performance on an immediately subsequent task is improved, impaired, or left unchanged by an initial task has a bearing on whether the capacity in question is a kind of knowledge, a skill, a constant capacity, or a depletable resource. More generally, it raises the question of how to develop a conception that accommodates Baumeister's findings.

This set of implications begins to answer the philosophical triviality objection. It shows that the ego depletion effect has a bearing on philosophical concerns about the nature of self-regulation and volition. I believe the philosophical implications are much broader and more striking. But to say more about them, I will first need to consider, and in a few cases accept, some other objections and potential revisions to Baumeister's strength model. This will put us in a better position to draw further philosophical conclusions. The last two sections of this chapter will be, in part, a fuller answer to the philosophical triviality objection.

(3) Conservation

Baumeister's conclusions can seem misguided in a different way. In the case of physical exercise or manual labor, it is unusual for a loss of strength or energy to force us to stop. Usually, we stop because we decide to stop, to save our energy for later. What appears to be depletion of a limited resource can actually be a motivated and strategic withholding of effort. In that case, the explanation for diminished performance can be not depletion, but deliberate conservation. We can call this *the conservation objection*.

I think this objection rests on a confusion, and can be resolved by more detailed attention to Baumeister's studies. Baumeister and his colleagues, particularly his student Mark Muraven, have themselves taken an increasing interest in conservation. Muraven (1998), for example, found performance on a second task to be improved by higher incentives such an increased monetary reward—and diminished by the expectation of a third task. In these cases, we perform better when the stakes are high, and less well with another task to anticipate. These results suggest conservation in response to an evaluation of one's own capacities and incentives, rather than an inability to persist or succeed. As Baumeister (2002, 133) puts it:

> The initial exercise does deplete the self's resources, not to a catastrophic degree, but enough to motivate the person to conserve what is left. This view would be most consistent with the analogy to a muscle. Athletes do not exert themselves at maximum output right up to the point of exhaustion. Rather, once their muscles begin to have fatigue, they conserve their energy. In the same way, the self might

be conserving its limited resources in case an urgent decision had to be made or a powerful influence needed to be stifled.

Conservation, in other words, is consistent both with depletion effects and with the muscle analogy. In a more recent article, Baumeister writes that "Ego depletion effects are mostly conservation effects rather than exhaustion effects"(2012, 113).

Ego depletion effects mostly *are* conservation effects. By the same token, the conservation effects in such cases are themselves ego depletion effects. As quoted earlier, ego depletion is "a temporary reduction in…capacity or willingness to engage in volitional action…caused by prior exercise of volition"(Baumeister, Bratslavsky, Muraven, and Tice 1998, 1253). Conservation might be most naturally thought of as a kind of unwillingness, rather than inability, to expend energy in a particular activity. The possibility of unwillingness does not suggest that there is not also genuine exhaustion, in which we are unable to act. On the contrary, unwillingness can depend on the possibility of exhaustion. We can see this by asking a simple question: why do we conserve in the first place? In the case of muscle strength, part of the answer can be that it is otherwise unpleasant or more difficult to engage in further activities.[122] But we also know that our muscles can fail, leaving us temporarily unable to perform physical tasks that are normally within our power. Here the muscle analogy is once again useful. If we decide to conserve our willpower, this can be at least partly because we believe that we will otherwise eventually become unable to exercise it even in circumstances in which we usually exercise it successfully.

In other words, the conservation objection backfires. To see ourselves *as conserving* is already to at least tend toward accepting the strength model. What is conserved in any conservation is normally a resource that is subject to depletion. We would not conserve if conserving did not improve our success on future tasks, or if those tasks did not use the same resource as the task on which we reduce our effort or persistence. Moreover, actual conservation can often depend on the thought that if we do not conserve, we will become exhausted, and be unable to go on. We may or may not have such a thought explicitly when conserving volitional strength. But conserving is itself a particular way of being unwilling to expend a limited resource. Conservation effects thus support both the strength model and the possibility of volitional incapacitation.

(*4) Self-fulfilling prophesies*

One might suspect that ego depletion effects are the result of a self-fulfilling prophesy. Our beliefs about our performance can in general affect our performance. And our beliefs about willpower or depletion can affect performance in dual task studies. Job, Dweck, and Walton (2010) suggest that people who *believe* the depletion model tend to do

[122] Even mere difficulty, without incapacity, can be understood as a strain on our capacity, impossible without at least the possibility of eventual incapacity to continue. But I do not insist on this point here.

less well, with lower success rates and lower duration of effort. A similar effect can be achieved with simple priming. In a study by Eric Miller and colleagues,

> Participants assigned to the limited resource theory group rated their agreement with items such as, 'Working on a strenuous mental task can make you feel tired such that you need a break before accomplishing a new task.' Participants assigned to the non-limited resource theory group rated their agreement with items such as, 'Sometimes, it is energizing to be fully absorbed with a demanding task.'[123]

Sure enough, the priming study found greater depletion in people primed to focus on their limitations.

These results are important, and I will come back to them. But they have little weight as a general objection to the strength model. Depletion can be affected by many factors, including a subject's beliefs, without removing the overall effect. People may be either unable or unwilling to engage in volitional activity when mistaken or primed, and a theory like Baumeister's can allow for this. That beliefs about willpower can affect the exercise of willpower is an important fact, but one that fills in the details of, rather than undermining, a conception of ego depletion.[124]

(5) The Scope of the Model

Baumeister's strength model is meant as a model of self-regulation, or "the capacity to alter or override one's responses, including thoughts, emotions, and actions"(Baumeister 2002, 129). But the notion of self-regulation can seem obscure in several ways, making it unclear what the model is meant to account for. First, one might think, surely self-regulation is not the mere *capacity* to alter or override one's responses, but the actual alteration or overriding. To have the capacity but not exercise it is to not regulate oneself. Second, altering and overriding can seem importantly different. When, for example, a desire is overridden, the desire can persist, but be resisted, most typically by

[123] Miller et al (2012, 1).

[124] In the text I do not consider the wide range of other alternative explanations, or the possibility of piecemeal alternative explanation for the full set of ego depletion effects. The answer to this sort of doubt is largely in the details of the studies, which are carefully constructed to eliminate alternative explanations. In the case of mood, for example, direct measures of mood found no differences between people who did and did not exercise self-regulation in a particular task (Muraven and Baumeister 2000, 253). Moreover, the mood regulation study used both positive and negative mood alteration, and found that the direction of mood regulation had no effect on performance in the second task. It was the regulation itself that had the depleting effect, even with an improvement in mood. Other studies controlled for negative affect (e.g., Baumeister, Bratslavsky, Muraven, and Tice 1998, Study 3), strength of impulses, demands on attention, and other factors. Though ruling out every possible alternative explanation would require a more detailed empirical treatment, the convergence of results from dozens of studies of a wide range of different tasks has no easy alternative explanation. For further discussion, see Muraven and Baumeister (2000, 252-3), and Hagger et al (2010, 498-9 and 517-8).

our choosing not to act on it. When a desire is altered, on the other hand, the desire no longer exists, or at least not in the same form. Altering and overriding can strike us as different phenomena, and we might want to know what they have in common that makes them both examples of self-regulation. Third, the notion of alteration can itself seem obscure. What is it to 'alter' one's own action? Is it to override a desire? Is it simply to act? If so, does acting in general count as self-regulation? Fourth, one might wonder: how conscious and how deliberate does self-regulation have to be? Understood narrowly, self-regulation is an essentially reflective and self-aware process. Understood broadly, it might include much more, though at some point it can become obscure in what sense a less self-aware process is 'self-regulation'.

Baumeister makes several related remarks about the notion of self-regulation in a series of articles, often in the context of a discussion of self-control. As he puts it (Baumeister 2012, 112), 'self-control'

> is largely synonymous with 'self-regulation', a term preferred by many researchers because of its greater precision. To regulate is to change: namely, to change in the direction of some standard, some idea about how something could or should be. Self-regulation thus means changing responses based on some rule, value, or ideal.

To regulate oneself is to change oneself based on some sort of standard. Though this can sound essentially deliberate or self-conscious, Baumeister believes it does not have to be. The full passage defining self-regulation (Baumeister 2002, 129) reads:

> The terms *self-regulation* and *self-control* refer to this capacity to alter or override one's responses, including thoughts, emotions, and actions. (In general, *self-regulation* is the broader term, encompassing both conscious and unconscious processes and sometimes referring to all behavior guided by goals or standards, whereas *self-control* refers more narrowly to conscious efforts to alter behavior, especially restraining impulses and resisting temptations. The distinction is not important in our work.)

Here it is clear that "self-regulation" encompasses unconscious processes as well.

On the other hand, it is not always clear in Baumeister's descriptions whether all action, or at least all intentional action, requires self-regulation. As he notes in the passage just quoted, "self-regulation" can "sometimes" refer to all behavior guided by goals or standards, and intentional action is often thought to essentially involve a goal, even when, in the limiting case, it is done for its own sake. Moreover, if giving rise to a response, or the response itself, is a kind of alteration of one's overall set of responses, then every action is an example of self-regulation. On the other hand, as Muraven and Baumeister (2000, 247) write, "Many behaviors (such as solving math problems) may be difficult and effortful but require minimal overriding or inhibiting of urges, behaviors, desires, or

emotions. Hence, not all effortful behaviors are self-control behaviors." By allowing that a difficult behavior can involve minimal overriding or inhibiting, Muraven and Baumeister implicitly distinguish difficulty or effort from the overriding of an urge or desire to stop. It might then be possible to engage in even a difficult or effortful intentional action without self-regulation or self-control. Still, even in this passage, one can wonder whether "minimal" overriding would be enough for self-control or self-regulation. As Hagger, Wood, and Stiff (2010, 499-500) note, this potential confusion about the demarcation of self-regulation is reflected in the ego depletion literature; difficult mathematics problems, for example, appear in several studies as the depleting self-regulation task, and in several others as the initial task in the nondepleted control group. As we saw, Baumeister himself includes anagram tasks but not mathematical problems as depleting tasks. It can be hard to see where to draw the line, and why to draw it at any particular task.

Though the demarcation of self-regulation is partly a terminological issue, Baumeister's claim of generality makes it significant. According to the third point in the strength model, quoted earlier, "all forms of self-regulation require some such resource, and indeed they may all draw on the same resource." Demarcating self-regulation too broadly risks including processes that do not draw on any sort of volitional resource. On the other hand, demarcating self-regulation more narrowly would not undermine the strength model. On the contrary: it would only leave open the possibility that the strength model applies to more than self-regulation. There is thus something to be said for a more restrictive conception of self-regulation, and a risk in beginning too broadly.[125]

Which processes do draw on a volitional resource is, of course, still to be determined, at least partly by the sort of studies Baumeister undertakes. So far, this is a reason to undertake more such studies, rather than to be suspicious of them. The extent to which a 'volitional activity' leaves us less able or willing to engage in *all* other such activities may still be the object of some suspicion, and may call for further study as well. But these doubts leave in place the central finding of an ego depletion effect across a wide variety of contexts.

On the other hand, doubts may persist with respect to the use of the term "self-regulation," along with its definition, to characterize Baumeister's findings. Apart from its use as a technical term in psychology, "self-regulation" may still suggest a self-aware, deliberative, and even computationally complex process. We may not yet have a clear enough picture of what is included under "alter or override," or why Baumeister writes of a mere capacity. I will avoid these issues in what follows, since my concern is the possibility and the basic characteristics of ego depletion phenomena, rather than the demarcation of

[125] Whether all intentional action requires self-regulation is a different question than the question of whether all intentional action involves ego depletion. Whether or not self-regulation is depleting, other activities may be depleting as well. As Hagger et al also note in the case of cognitively difficult mathematical problems, depletion by effort in general would only widen the range of application of the depletion model.

I am inclined to think that all intentional action gives rise to at least some, even if negligible, ego depletion. But since my aim here is only to introduce a notion of executive fatigue to account for a particular kind of inaction, I do not take a stand on this issue in the text.

their scope. This is why, when I draw on Baumeister's findings in defending the Identity View in the next section, I will avoid the notion of self-regulation.

(6) The Nature of the Resource

I have not yet considered a question that cuts to the heart of Baumeister's strength model: What exactly is depleted? Not only the scope but the nature of the explanation can still seem obscure.

It might be unfair to expect Baumeister to offer a fully developed conception of the volitional resource he has in mind. Ego depletion studies can be seen as contributing to the development of such a conception, while at the same time showing that we need one. An initial task's effect on performance in a seemingly unrelated, subsequent task suggests the presence of an underlying resource drawn on by both tasks. So although there is more to learn about the underlying resource, Baumeister might say, the ego depletion studies offer a crucial pointer in the right direction.

Baumeister, Bratslavsky, Muraven, and Tice (1998, 1253) suggest such a view in a brief discussion of Freud: "Freud was rather vague and inconsistent about where the ego's energy came from, but he recognized the conceptual value of postulating that the ego operated on an energy model." Vagueness and inconsistency about the source—and, we can add, about the nature—of the energy might not undermine the usefulness of thinking in terms of energy that is drawn on and used in various tasks.

Nevertheless, there is a remaining problem about *any* use of the notion of a resource in thinking about willpower. It can seem unclear whether the failure to use a resource is due to a lack of the resource, or a failure to *use* the resource that one has. If the resource is itself a kind of volitional energy, it can seem necessarily unclear how to draw this distinction. One might ask: does the subject lack motivation, or lack motivation to use the motivation? What would it be to have willpower, but be unable to use it? The difference between having and using a resource in this context is difficult to make out.[126]

[126] For versions of this point, see Navon (1984), and Hagger, Wood, and Stiff (2010, 515). As these earlier discussions suggest, difficulties concerning the notion of a resource make claims about the resource difficult to test empirically without assuming a resource-based interpretation of one's findings. Navon cautions that a resource-based theory can be "self-reinforcing" and "unfalsifiable"(1984, 231), and thus tends to be "unparsimonious" and "not more explanatory" than other theories. I think these relatively technical difficulties can be made more intuitive by noticing Baumeister's uneasy combination of metaphors: on the one hand, the prominent use of 'ego', suggesting that *I* am depleted; and on the other hand, the image of a reservoir of energy, suggesting a kind of lake of willpower in one's mind. Here it is already unclear whether what is depleted is a person, or a tool which is external to the person and can be used at will.

At this point the notion of willpower can seem especially dubious, and the strength model incoherent. If there is no justification for talking of a resource, there seems to be no justification for talking of its depletion or restoration.

In what follows, I will avoid the notion of a resource, and the image of willpower as a reservoir of stored energy. Instead, I will consider the 'power' in 'willpower' simply as a kind of ability. Most importantly for our purposes, many of the phenomena of 'ego depletion' are examples of diminished ability to execute one's intentions in the face of temptation, reluctance, or other psychological obstacles. Our volitional capacities might not be measurable by an ego-meter. But they can still be significantly exhausted.

IV. Executive Fatigue

1. Intending while Tired

When one comes home after a long day, one comes with many intentions. There is a project to finish, a friendship to repair, an oppression to fight, a phone call to make, and a dinner to cook. But sometimes, what one does is collapse on the couch. The collapsing can itself be intentional, and long awaited. But it might not be. One can be passing the couch on the way to the computer, and suddenly find oneself lying on it. If someone asks how one ended up on the couch, it is easy to come up with an explanation. One can just say: "I was tired."

We usually do not hesitate to attribute beliefs to people who are asleep or unconscious. When asking someone whether her husband believes in God, or believes that abortion is wrong, or believes that a Democrat will win the next presidential election, we would find it absurd to be told: "Hold on, let me check if he's sleeping." A temporary inability to consciously express or reason from a belief does not stop us from attributing the belief.

In the same way, we continue to attribute intentions to sleeping or unconscious people. We can sit near a sleeping relative and tell someone about her intentions. We say: "My daughter will call you today," as a report of her intention; or, "She fully intends to propose marriage this summer"; or, "She really is in a three-year program intentionally," even as we see that she is entirely unable to act on the intention at the time. Although sleep is a relatively simple example in this context, we can do the same with someone who is temporarily malnourished, or too tired after a long day to even think about writing, marriage, or college. And as these examples suggest, the intentions in question can be both intentions for the future, and intentions for currently ongoing pursuits.

These examples of tiredness and sleep illustrate two general and uncontroversial ideas. One is that there is a familiar distinction between having an intention and being able to act on it in a particular moment. The other is that what prevents someone from carrying out an intention can be a state of the person, rather than a purely 'external' obstacle such as a broken phone, a flight delay, or a lack of money. Sometimes we ourselves are too tired,

or not awake enough, to take any steps toward the realization of our intentions, goals, or plans.

Fatigue can account for failure to act, while allowing that we still intend to act. It allows a distinctive kind of failure that can explain the kind of lack of motivation with which this chapter started, while still justifying the attribution of an intention. I will begin by characterizing the kind of fatigue I have in mind in more detail, and then go on to explain its importance for a defense of the Identity View.

Tiredness comes in many forms. We can be moderately tired in a way that has little impact on our activity; we can be exhausted and barely able to continue; in more extreme cases, we can literally collapse, and be unable to move at all.[127] I will use the word "fatigue" loosely, to refer to a relatively severe form of tiredness. I make no assumption that a person is either able or unable to act while fatigued, leaving open the possibility both of overcoming fatigue and of being rendered unable to act by it.

We recognize a relatively simple form of fatigue in the use of our muscles. Muscle fatigue is recognizable, for example, in muscle failure after heavy lifting. We report that our muscles feel tired, and we find it more and more difficult to use them. When suffering from malnutrition, or from a neurological disorder, we can experience fatigue in all the muscles of our body. But the fatigue can also be quite specific; after too many bicep curls, walking can be as easy as ever, but lifting can be frustratingly difficult. Once again, fatigue may not require inability to go on. Even when lifting heavy weights, we often stop before physical muscle failure, to conserve our energy or avoid greater fatigue. But at some point even electricity will not induce muscle activation.

Other forms of fatigue can be specifically cognitive. When reading or writing a difficult text, or solving a series of math problems, we can find it increasingly difficult to think. We report feeling tired, and find it more and more difficult to go on. We eventually stop to do something easier—such as, for example, vigorous physical exercise. Going on a run can be a relaxing break from a mentally exhausting activity.

There is another form of fatigue, which can be called "executive fatigue." Its primary locus is neither muscle activation nor thought, but the execution of intentions. It is a fatigue specific to the will: a volitional analogue of muscular and cognitive fatigue. We face it when battling a powerful addiction, or when restraining our anger toward a close person in our lives. We face it even in comparatively easy activities such as the ones studied by Baumeister and his colleagues: suppressing thoughts of white bears, resisting chocolates, and solving anagrams.

Volitional exertion can at times leave us with an impaired ability to carry out or execute our intentions. We often deliberately conserve our energy to avoid such situations. But our giving up on volitionally demanding tasks is not all calculating or deliberate. Extended volitional activity, such as a long day of shopping, or studying for finals, or

[127] These and other differences are sometimes described using a distinction between tiredness, fatigue, and exhaustion, especially in the context of nursing. Since the relative extent or extremity of fatigue in not important for my purposes, I leave out these distinctions here; but see, for example, Olson (2007).

suppressing intrusive thoughts, can leave us vulnerable to temptation and to various forms of irrationality. This much is not controversial. But it is easy to miss these facts if one dismisses all talk of willpower, and it is easy to miss their significance. The basic point is this: we can continue to have an intention, even in executive fatigue. We can simply be too volitionally exhausted to do what we intend to do.

Executive fatigue is weakness of will, in one sense of that phrase.[128] It is the inability to carry out what one wills, or at least what one intends. In many cases, the weakness is temporary or intermittent. One might experience it only during finals, or on an occasional long day of shopping. But there can be also be a volitional analogue of chronic fatigue, in which one's ability to execute intentions drops dramatically and for an extended length of time. Such fatigue can begin after an extremely demanding ordeal, such as a divorce or a cancer treatment, or perhaps, in some cases, without a discernible explanation. There might also be people who are weak willed in this sense throughout their lives. For them, the weakness might be more naturally seen as a trait of character, rather than an ongoing condition of fatigue. I will not try to decide this issue here.

Still, it is worth noting a potentially confusing terminological overlap. *Akrasia* is often called "weakness of will." In philosophical contexts, one often calls someone who does or intends something she believes she should not do "weak-willed." This can be a harmless technical or stipulative use, though it can also bias our attention to some examples rather than others. But calling executive fatigue "weakness of will" can make it seem as if I am assimilating lack of motivation to *akrasia*. One might think: Executive fatigue is weakness of will; *akrasia* is weakness of will; so executive fatigue is *akrasia*. This would be confusing an ordinary notion with a technical one. Executive fatigue, I suggest, is weakness of will in roughly the ordinary sense. It is a diminished capacity to execute intentions that one does have, rather than having an intention to do what one believes one should not. In this case—to go along temporarily with talk of weakness of will—we do will something, but we do not have the strength to carry it out.

This fatigue offers a way to account for the counterexamples with which I began this chapter. According to the Identity View, an intention is a normative belief. In cases of lack of motivation, one can seem to believe one ought to get up, make a call, or donate to charity, without any corresponding intention. If we did intend, it seems, our intention should be effective in these cases. But with nothing to show for it, it can seem unclear why one should think there is an intention there at all.

Executive fatigue offers a way to account for failing to act on a normative belief, without lacking the intention. One can simply be tired, or volitionally exhausted, as we sometimes are with respect to many of our intentions. There is thus an alternative to denying that the intention is present. The failure can be not in the intention, but in its execution.

Once we see that this sort of failure is possible, we can recognize it in an even broader range of cases than Baumeister considers. If Baumeister is right, all volitionally

[128] For discussion of this use of "weakness of will," see Holton (2009), esp. Chapter 4.

demanding tasks deplete a single resource or capacity. So on Baumeister's model, it can seem as if non-akratic lack of motivation must always be global, rather than specific to, for example, making a phone call. This limitation fits naturally with the status of failure to get out of bed as the most common paradigm case of lack of motivation. In such cases, one sometimes says, one is not motivated to do *anything*. With local phenomena such as making a phone call or donating to Oxfam, there can also be a more localized failure. The failure might be described as fatigue; calling one's cousin might be especially taxing and, one might say, "I'm so tired of doing it." Or in some cases we might speak more naturally of the obstacle to be overcome, rather than focusing on the diminished capacity to overcome it. It might be fear that keeps me from making the call, or greed that keeps me from executing my intention to donate to Oxfam. Neither the obstacle nor the failure to overcome it require the presence of *akrasia*. Fear or greed can distract us, or simply be too much for us to successfully move past, even when we intend to. Such local failure would be more like a 'glitch' that leaves us unable to execute a particular intention, rather than an overall depletion of our general volitional or executive capacity. Once we recognize the possibility of executive failure, we can also recognize failures specific to particular activities. Our general volitional capacities can be diminished to the point that certain, especially volitionally demanding tasks are too much for us.

Indeed, many different explanations of the apparent failure can be appropriate. As a general explanatory strategy, we can combine the four explanations I have considered of apparent failure to act on our normative beliefs. First (§II.1), the beliefs themselves can sometimes be merely apparent. We can stay in bed, believing only that we "should", according to social convention, be up by now. Our actual normative beliefs might favor resting, or they might allow either staying in bed or getting up. Second (§II.2), we can akratically refrain. We might intend, against our own better judgment, not to get up, even though we believe we ought to. I considered these cases in Chapter 4. Third (§II.3), the normative belief can be conditional. We might believe that we should get out of bed if there is something interesting to do, but that, for now, that condition is not satisfied. Fourth (§§III-V), we can be too tired. Our capacity to overcome resistance in executing our intentions can be challenged to a degree that makes it difficult or even impossible to get up, even when we intend to and do not intend not to. In some cases we might be intentionally conserving our energy, but in other cases we might not be. Without ever deciding against it, we might just fail to get out of bed.

Each of these four explanations can be correct in some cases. It is in general possible to have related non-normative beliefs, to refrain akratically, to have merely conditional normative beliefs, or to be temporarily too tired to execute an intention. And any of these can make it look as though we do not really intend to do what we believe we ought to. We can sincerely say, "I should get out of bed now," and do nothing, while the implicit condition, inverted commas, akratic intention, or executive fatigue remain unobvious. In each case, we can seem to have a normative belief but lack the corresponding intention. In each case, the explanation can explain the appearance while also explaining why it is a mere appearance.

There is, then, no need to choose a single, general explanation from among the four. They are much more powerful together. Nor have I ruled out the possibility of other, similar explanations. There might be other phenomena which at least sometimes create an appearance of normative belief without intention. And there might be psychologically and philosophically significant disagreement about which explanation applies in a particular case. Discouragement, for example, might include any one or more of the four. When we fail to donate to charity, or when we are depressed or psychopathic, the explanation can again be one of the four I considered, or a combination of them. Depending on the case, one might not genuinely believe one ought to act; or one might have a conflicting intention to keep one's money; or one's belief might be a complex or conditional one; or one can intend to donate, but not be able to 'work up' the energy. The variety of explanations helps capture some of the subtleties of our inaction.

Most importantly, there are no clear cases of normative belief in which an apparent lack of a corresponding intention must be attributed to an actual lack of intention. On this point, executive fatigue plays a particularly important role. Unlike the other three explanations I considered, the possibility of executive fatigue severs the apparent necessary connection between having an intention, with no external obstacles and no conflicting intentions, and acting on that intention. There is always the possibility of, as we ordinarily say, lacking the willpower to do what we intend to. Executive fatigue thus plays a central role in explaining genuine lack of motivation as a distinctive counterexample to the Identity View. It is how we can non-akratically fail to do what we do believe we ought to.

V. Implications

As in earlier chapters, I have not argued in general terms that intention is best understood as a normative belief. I will defend that view in Chapter 7. Instead, this chapter defends two more modest ideas particular to one kind of counterexample. First, we have not seen a case of lack of motivation that provides a compelling counterexample to the Identity View. There is no failure to act on a normative belief for which the Identity View cannot offer an explanation. Second, the Identity View can help shed light on the details of our own activity. To further defend this second idea, I will conclude by considering some of the broader implications of this chapter.

The possibility of executive fatigue—or, in Baumeister's terms, ego depletion or loss of willpower—suggests that, to understand our everyday activity, we need the notion of an executive capacity. This capacity to execute our intentions can be diminished in a way that is distinct from diminished access to means such as tools, bodily mobility, or our own intelligence. And it can explain our failure to do what we believe we ought to do, while allowing that we intend to do it.

To be sure, we need more investigation into the nature of this capacity. Some of it will be empirical. Psychologists will continue to investigate the range of 'ego depletion'; the possibility of its specificity to particular domains; and the significance of deliberate

energy conservation, beliefs and priming about willpower and related topics, and 'bottom-up' processes such as increases in blood glucose. Conceptual investigation is needed too. It is not yet clear to what extent it can make sense to talk of a resource, or of a muscle, or to what extent a failure of willpower can be a failure of rationality. But whatever one's views on these issues, there is a general conclusion to be drawn. Lack of motivation need not be lack of intention. It can be a failure to execute the intention, rather than a failure to have it.

The phenomenon of executive fatigue also has practical implications which are both plausible and interesting, concerning the attitudes we take to our own willpower or executive capacity. We should, in general, be aware that our activity can deplete our volitional capacities, leaving us drained or volitionally fatigued and less able to perform well. We can avoid situations that drain our willpower, especially when we know we will soon need it. As one discussion of Baumeister puts it (Aamodt and Wang, 2008):

> In the short term, you should spend your limited willpower budget wisely. For example, if you do not want to drink too much at a party, then on the way to the festivities, you should not deplete your willpower by window shopping for items you cannot afford….On the other hand, if you need to study for a big exam, it might be smart to let the housecleaning slide to conserve your willpower for the more important job. Similarly, it can be counterproductive to work toward multiple goals at the same time if your willpower cannot cover all the efforts that are required.

Baumeister and Tierney (2011, 38), for example, recommend making only one New Year's resolution, citing evidence that people who make several are less like to keep even one. In the longer term, Baumeister's findings suggest that three factors counteract ego depletion: gradual buildup through practice of demanding tasks; rest; and self-affirmation, specifically of our capacity to successfully use willpower.[129]

If this were the whole story, it would be, ethically speaking, relatively straightforward. But in fact the situation is much more complex. Recall the self-fulfilling prophesy objection (§III, objection 4). People who *believe* that their willpower is significantly limited, or are primed with phrases such as "Working on a strenuous mental task can make you feel tired such that you need a break before accomplishing a new task," do *less* well on willpower tasks. Awareness of the limitations of willpower is a double-edged sword. Job et al (2010, 1692) write: "People who learn about the strength model of self-control may conclude that they are at the mercy of a fixed, physiological process that limits their willpower." As the priming study suggests, no full-blown conclusion or belief is necessary. Even a reminder of the possibility of fatigue can impair performance. It might, of course, do this partly through deliberate conservation. But we should also be aware of the potentially discouraging effects of reminding oneself of the weakness of one's will. There is, in other words, the possibility both of excessive arrogance and of excessive

[129] See, for example, Baumeister and Exline (1999) and Baumeister (2012).

humility. One can dismiss or underestimate the empirical limitations on one's volitional capacities. And one can harp on them, in a way that can make one become tired or discouraged and accomplish less. Taken together, the empirical studies of willpower raise an ethical challenge: How can we find the right way to take our limited willpower into account, without overemphasizing it?

This challenge echoes a familiar problem about freedom. According to Kant, a being with a will "cannot act otherwise than under the idea of freedom"(1998, 448).[130] To act at all is, in part, to regard oneself as undetermined by outside causes. But the natural world, ourselves included, seems governed by empirical laws. "There arises," Kant says, "a dialectic of reason, since the freedom attributed to the will seems to contradict the necessity of nature"(1998, 455). We are faced with the problem of reconciling these two views of ourselves: the view of ourselves as free, which we must take up when we act, and the view of ourselves as part of the order of nature.

In the case of willpower, the problem is not a conflict between theoretical and practical points of view, or even a threat of determinism. We can both over- and underestimate our capacity to succeed in demanding tasks, taken as a purely empirical matter. And we can err in either direction practically. Out of laziness, lack of self-respect, or unwillingness to take responsibility for our actions, we can constantly remind ourselves of our failures, or stingily conserve willpower for unspecified later projects that we never take on. We can also emphasize our capacity to overcome apparent limitations on our willpower in a way that makes us reckless or arrogant. So in one way, the problem is more complex than Kant's. It is a problem of balancing practical considerations both with empirical ones and with each other.

Consider a sports coach at a game. A pep talk to the team can be filled with wildly unlikely statements about the draining task faced by the players, especially when they are tired and losing. "We can get through this!" "We're the best team that's ever lived!" "Nothing can stop us!" The coach usually knows these statements are not all literally true, and often the teammates know it. But it is not an accident that they keep being made. As an attempt to work up determination, they are often successful. They succeed, partly through their effect on the players' view of their willpower. They prime the team to expect success, and might in some cases, at least temporarily, lead the players to see their willpower as unlimited. The coach makes the statements out of practical considerations; his goal is not to describe but to win. He wants the team's determination to be increased and effective. His result is not only a means to winning, but also, for some people, the most exhilarating moment of play: a sense of one's own freedom, a sense that one can do anything.

In other contexts, accurate self-assessment takes pride of place. Seeing that, every New Year, one makes five resolutions and keeps none, one may come to just make one and keep it. An understanding of one's volitional limitations requires a degree of modesty both in one's self-estimation, and in the number and difficulty of projects or temptations one takes on. We should not spread ourselves so thin that we exhaust ourselves without getting

[130] As in earlier chapters, I use the standard Akademie pagination of Kant's writings, found in the margins of most editions.

anywhere. On the other hand, as the priming study suggests, reminding oneself of the extent of one's volitional capacity can be effective without being misleading. This suggests a practice of keeping oneself primed: that is, keeping oneself reminded of one's abilities and successes to maintain an attitude of confidence. That attitude can be both a virtue in itself, and an effective tool.

Indeed, there seem to be two contrasting virtues or norms here. An understanding of willpower calls for both a kind of modesty, and a kind of confidence. A healthy modesty will limit how much volitional strain we take on, especially in planning or in accepting new challenges. Of course, it is possible to be too modest, and take on too little. But as my earlier consideration of the triviality objection already suggests, we often systematically underestimate the strain that even relatively easy tasks will place on us. Modesty reminds us to limit our exposure to volitional strain, protecting ourselves from a fatigue that can leave us dazed, unproductive, and unhappy.[131]

Confidence is required especially in execution. We should set up our lives to be manageable, but, like a sports team, treat ourselves as unstoppable in the execution of our goals. Once we decide to only make one New Year's resolution, we should, in most contexts, ignore the fact that we would not have kept the other four. Though we must in some contexts take our volitional limitations into account, it is also a good idea to limit our attention to and experience of those limitations. While limiting our exposure to volitional strain, in other words, we should focus our attention on our ability to overcome it.[132]

---

[131] Many empirical studies have found what Lowenstein (1996) called a "cold-to-hot empathy gap," or, as Nordgren, van Harreveld, and van der Pligt (2009, 1523) put it, "a restraint bias: a tendency for people to overestimate their capacity for impulse control." Lowenstein points to our limited memory for visceral experience to explain the 'gap'; but whatever the explanation, we seem to have a powerful tendency toward immodesty in our estimation of how much volitional resistance we can handle. See Nordgren, van Harreveld, and van der Pligt (2009) for discussion of more recent work.

[132] I attempt to stay neutral here on whether to think of modesty and confidence in terms of virtues or of imperative-like norms. But I follow Aristotle in choosing a name for the good or virtuous state that contrasts more strongly with the extreme we typically tend toward. According to Aristotle (1999, 1109a1-18):

> In some cases the deficiency, in others the excess, is more opposed to the intermediate condition. For instance, cowardice, the deficiency, not rashness, the excess, is more opposed to bravery, whereas intemperance, the excess, not insensibility, the deficiency, is more opposed to temperance.
>
> This happens for two reasons: One reason is derived from the object itself. Since sometimes one extreme is closer and more similar to the intermediate condition, we oppose the contrary extreme….The other reason is derived from ourselves. For when we ourselves have some natural tendency to one extreme more than to the other, this extreme appears more opposed to the intermediate condition. Since, for instance, we have more of a natural tendency to pleasure, we drift more easily toward intemperance than toward orderliness. Hence we say that an extreme is more contrary if we naturally develop more in that direction.

These virtues or norms of modesty and confidence offer a way to systematize popular wisdom about willpower. It is modesty that can ask us to rest, eat healthy, stay close to other people and ask for help, form specific implementation intentions, and set up an irreversible reward or threat when needed. These are ways of complementing or making up for our essentially limited abilities. But it is part of confidence to keep one's mind on what one most wants, emphasize one's successes, and refuse to harp on or blame oneself for one's failures. We can also see more clearly how these bits of popular wisdom or advice can interact. Modesty about our limitations can prompt us to take measures to boost our confidence.

Aiming for a balance of modesty and confidence begins to meet the challenge of finding the right attitude to one's own willpower. But the challenge is still complex. It is partly a question of simple empirical accuracy: of having an undistorted picture of how much work or family one can handle in a week, how much shopping or studying one can withstand, and how many New Year's resolutions one can succeed in keeping. The challenge is, secondly, one of finding the proper attitude to one's own freedom. Third, it is a challenge of finding an attitude that will help one achieve one's ends. And fourth, it is the challenge of balancing these three kinds of consideration: the demands of accuracy, the awareness of freedom, and the attaining of intended effects.[133] These intricately interrelated demands become all the more complex when considering the willpower of others, to whom one can easily become unsupportive or paternalistic. If I am right, these issues should be treated as paradigmatically philosophical, rather than relegated to the quagmires of self-help. I have suggested two guidelines for meeting the challenge: modesty in how much one takes on, and a practical confidence that one can do what one sets one's mind to. There is more to say about the nature of these attitudes, and about how to understand and meet their demands without sliding into a life that is too small or too arrogant. But the challenges are recognizable, and useful to articulate.

Willpower can seem like a conceptual swampland: an obscure kind of power in an obscure kind of will. I have argued that there is a relatively straightforward way of understanding it, as our capacity to persist in the execution of our intentions. There is no need to appeal to an obscure sort of will, or an intangible reservoir of stored energy. But if I am right, willpower is an *ethical* swampland. Modesty and confidence, the two central

---

Just as we tend toward intemperance rather than insensibility, we seem to tend toward immodesty rather than overmodesty in exposing ourselves to strains on our volitional capacities. On the other hand, it might be that we tend toward discouragement or low self-confidence rather than brashness in when it comes to persevering in the face of difficulties in execution. I choose the terms 'modesty' and 'confidence' in response to these tendencies, as well as to highlight their interrelation and potential tension with each other. But I do not argue these points in detail here, and not much depends on the names.

[133] Baumeister (1989) suggests in a somewhat different context that there is an "optimal margin of illusion," in which a confidence that goes moderately beyond accuracy allows us to be most effective. The possibility of such a margin is one way of illustrating the tension, raising the question whether maintaining oneself in a state of moderate illusion is impermissible, troublingly self-deceptive, or simply prudent.

responses I recommended, easily come into tension with each other. Setting proper limits for our tasks can so easily lead us to become discouraged about our capacity to overcome those limits. And as we learn in sports, or in love, or in war, focusing on our capacity to overcome our limitations can so easily lead us to take on too many challenges and undermine our chances of success. This tension helps explain why success with willpower takes so much practice, both in self-discipline and in setting up one's life to avoid overcommitment. The other tension, between the demands of accuracy, proper awareness of freedom, and achieving intended effects, compounds the problem. Apart from its more direct implications for the Identity View, an understanding of executive fatigue helps express these difficulties, and begins to lead toward a solution to them. In this way, the Identity View itself leads us to understand and appreciate features of ourselves that we might otherwise ignore.

On the Identity View, an intention is a normative belief. When we fail to act on our normative beliefs, the Identity View cannot allow that our intentions simply do not match those beliefs. It is forced to see us as failing to act on our intentions. This can seem to be an unfortunate conclusion. But as I have tried to show, the conclusion is not unfortunate. We have not yet seen a compelling counterexample, in which someone has a normative belief but no corresponding intention. And we have a way of understanding failure to act on a normative belief, in which the failure is a failure in execution, rather than a conflict between belief and intention. Even when we are not akratic, we can fail in this way. This is a recognizable kind of failure. At the same time, an understanding of it sheds light on our everyday activities and attitudes. As we saw in considering Baumeister's studies, the extent of our failure is consistently surprising. As I suggested in this section, an understanding of executive fatigue has fairly intricate practical implications, and raises further questions that we need to ask. Once again, the Identity View is both defensible against apparently compelling counterexamples, and illuminating for an understanding of action and intention.

With this chapter I conclude the treatment of counterexamples, in which either normative belief or intention seems lacking or the two appear to be in conflict. In the next chapter, I turn to a more general defense of the Identity View as a conception of intention.

Chapter 7: Intention as Normative Belief

I. Toward a General Theory

I intend to return a book to the library. What is the difference between intending to return the book, and believing I ought to return it? In the past five chapters, I have defended the idea that there is no difference. An intention is a belief that one ought. Or as I put it earlier: A's intention to *x* is a belief that A ought to *x* (where this belief is first-personal, of the form: I ought to x). To give it a name, I called this view *The Identity View*. This ambitious view faces a range of apparently powerful counterexamples. But as I have argued, none of these counterexamples are decisive. The Identity View can address all of them, while at the same time shedding light on the details of our actions and intentions in a variety of circumstances.

Although it can be stated in one short sentence, the Identity View is the heart of a general theory of intention. It offers a way of understanding what intention is: an intention is a particular kind of belief. This chapter begins to spell out the theory, considering a set of more general issues about the nature of intention and belief. I will not consider competing theories here, let alone argue that they are inferior. The aim is positive: to spell out what thinking of intention as normative belief entails, and why the view is believable.

A theoretical conception of intention faces many questions, not all of which take the form of counterexamples. I will not be able to consider all of them. This chapter will have little to say about, for example, how intention gives rise to action. I make no claim to offer a fully developed theory. Instead, having considered counterexamples to guise-of-the-good views in earlier chapters, I want to *begin* the larger theoretical project, by sketching the outlines of a general theory of intention. The aim of this chapter is to explain how an intention could *be* a belief at all.

To do this, I want to consider a series of more abstract challenges to the Identity View. Intention and belief can seem to be fundamentally different kinds of state. But there is a variety of apparent fundamental differences, just as there is a variety of apparent counterexamples. Beliefs, one might think, aim to match the world, while intentions aim to make the world match them. Beliefs can seem not to be under our voluntary control, the way actions and intentions are. Beliefs and intentions can seem bound up with importantly different kinds of reasoning. As we will see, these objections are ways of spelling out the thought that intention is a fundamentally *practical* state, in a way belief is not. The Identity View can seem unable to do justice to this thought—and thus unable to capture what is distinctive about intention.

The guiding idea of this chapter is that a belief that I ought to act a certain way *is* a 'practical' state. When I believe I ought to return a book, this belief is itself 'practical', in the ways intention is practical. This is not just a point about the uses of the word

'practical'. Belief can be and do what intention is and does. What is true of my intention to return the book is true of my belief that I ought to return the book. Conversely, what is not true of my belief that I ought to return the book is not true of my intention to return the book, either.

Although this chapter focuses on theoretical concerns, rather than on examples, it has the same double aim as the previous chapters. As before, I will try to defend the Identity View against a series of objections, while also using those objections to develop the view in more detail. I argue, in §II, that a belief that one ought to do something, and an intention to do it, both 'aim to fit' whether one ought to—and both 'aim' to make one's actions 'fit' them. I argue in §III that, although we at least normally do not have voluntary control over our beliefs, we do not have that kind of control over our intentions, either. In §IV, I argue that practical reasoning can be understood as a species of 'theoretical' reasoning, or reasoning to and from belief.

If I am right, intention *is* fundamentally different from most of our ordinary beliefs, like the belief that it will rain, or that today is Sunday. But this difference does not show that intention is not a species of belief. What it shows is that the species is a very distinctive one. The belief that I ought to return a book is itself importantly different from other kinds of belief. It is a belief I can act on, and reach by practical reasoning. Intention is a particular, highly distinctive, genuinely 'practical' species of belief.[134] To defend this view, I turn now to considering the theoretical challenges to it.

II. Direction of Fit

Suppose you believe it is raining. But then you look outside, and see that the sky has cleared. In normal conditions, you change your mind, giving up your belief. Faced with a mismatch between your belief and the weather, you *could* instead try to change the weather to accord with your belief that it is raining. You could try it; but it would be strange. Here "strange" can be understood both statistically and evaluatively. It is unusual to try to change the world to match one's beliefs in this way. And it also seems inappropriate.

Intention seems very different. You might intend to be returning a book to the library, but get distracted, and start walking out without returning it. If you notice that you are not returning the book, you do not usually just drop your intention. Instead, you alter your movements and return the book. In this case, changing one's mental state is itself

[134] Compare Velleman (1989, 11), introducing his view that intentions are a kind of predictive belief: "The peculiarities of intention can best be appreciated as those by which a particular kind of belief would be set apart from other beliefs." I would say, as Velleman does here, that "My modeling intention on belief, and deliberation on reflective theorizing, is not an attempt to eliminate the practical as a distinct category but to make the category stand out." I doubt Velleman's ingenious defenses of his own view succeed; but Velleman's defense is complex, and I do not go on to consider it here.

what seems unusual and inappropriate. Faced with an intention that is not being realized, we do not usually change our minds; we change the world.

This contrast is often described as a difference in "direction of fit." Our beliefs, it seems, aim to fit the way the world is. Our intentions aim to make the world fit them. The uses of "aim" and "fit" here are metaphorical, and can be replaced with others or made precise in various ways. But however one describes the details, the sense of basic contrast is powerful. And if intentions and beliefs have different directions of fit, does this not show that intentions cannot be beliefs? We can call this *the direction-of-fit objection*.

The notion of direction of fit is somewhat obscure, and some powerful doubts have been raised about it. As we have already seen, there is no one clear way of understanding the central notion. On the one hand, it might be understood purely descriptively. Smith (1994, 115) offers one classic descriptive conception:

> The difference between beliefs and desires in terms of direction of fit can be seen to amount to a difference in the functional roles of belief and desire. Very roughly, and simplifying somewhat, it amounts, *inter alia*, to a difference in the counterfactual dependence of a belief that *p* and a desire that *p* on a perception with the content that not *p*: a belief that *p* tends to go out of existence in the presence of a perception with the content that not *p*, whereas a desire that p tends to endure, disposing the subject in that state to bring it about that *p*. Thus, we might say, attributions of beliefs and desires require that different kinds of counterfactuals are true of the subject to whom they are attributed.

On Smith's dispositional view, when I perceive that I am not returning my book to the library, my belief that I am returning my book "tends to go out of existence," while my desire or intention 'that I am returning my book' does not. In contrast to this purely descriptive conception of direction of fit, Searle (1983, 7–8) writes: "The idea of direction of fit is that of responsibility for fitting….It is, so to speak, the fault of the world if it fails to match the intention or the desire." Or as Platts (1979, 256–57) puts it: "The distinction is in terms of the direction of fit of mental states with the world…..The world, crudely, should be changed to fit our desires, not vice versa."[135]

Three kinds of doubt can be raised here. First, talk of 'direction of fit' seems disunified, since the central notion is ambiguous between descriptive and normative interpretations. Second, each of these interpretations might be further subdivided; we have not seen why Smith's way of understanding direction of fit should be the only available descriptive one, or Searle's the only available normative one. Third, for each conception of direction of fit, one might wonder whether it is true, or even coherent. Stubborn or wishful beliefs might not go out of existence upon the perception that not *p*; sadistic desires may not be ones that the world 'ought' to fit. Indeed, one might wonder whether there are any states that the world 'ought' to fit, just because someone is in that state; or in what sense

[135] For more recent normative conceptions of direction of fit, see Zangwill (1998) and Sherkoske (2010).

the world 'ought' to fit them; or what it means for the world to have a 'responsibility' or be at fault. It is far from clear that a coherent contrast can be drawn here at all, let alone used to describe a difference between belief and intention.

These criticisms have been made before, and I sympathize with many of them.[136] But I think the direction-of-fit objection retains some intuitive force, even if we accept that the notion of direction of fit is too ambiguous or too troubled to be of much use. Even a systematic classification and attack on conceptions of direction of fit might not completely dispel the sense that there is some important contrast to be drawn here between belief and intention. And although the objection can seem initially very powerful, I think it is not hard to answer, even if we accept the terminology of direction of fit. So instead of pressing concerns about the notion of direction of fit, I think we can allow its use, and more directly dispel the sense of disanalogy created by the direction-of-fit objection.

A belief that I ought to return a book has belief's typical direction of fit. I might be sick, or held up at gunpoint with the book in my bag; or a friend on her way to the library might offer to return the book for me. I can then conclude that it is no longer true that I ought to return the book. I then normally give up the belief. I would not refuse an offer to return the book for me, just to keep my belief that I ought to return it myself.

But my belief that I ought to return the book also has a regular and appropriate pattern of interaction with my actually returning the book. I might, of course, believe I ought to return the book, and still intentionally keep it; we saw such 'akratic' cases in Chapter 4. But normally, if I believe I ought to return the book, I do return it. And my returning it seems appropriately connected to the belief that I ought to. It is normal and appropriate to, as we often say, act on that belief. With respect to whether I ought to return the book, my beliefs aim to fit how things are; but in actually returning the book, I aim to make things as I believe they ought to be.

Intention is naturally understood in an analogous way. If I intend to return the book, my actually returning the book is a way I alter the world to accord with my intention, rather than changing my intention to accord with whether I will or am or did return the book. But I also tend to, and appropriately do, change my intention to accord with whether I ought to return it. If I am faced with an illness, or a gun, or an offer of help, I normally conclude that it is not true that I ought to return the book. And I normally change my intention. Like belief about what we ought to do, intention stands in two relations. We try to change the world to be as we intend it to be; and we try to adapt our intentions to how we ought to act. Both intentions and normative beliefs, in other words, can be said to 'aim to match' what we ought to do, and to aim to make the world match them.

Here one might object that the terminology of direction of fit can be applied in different ways. We can say that the normative belief has two directions of fit—one with respect to what one ought to do, and another with respect to what one actually does—and

[136] For criticisms of the notion of direction of fit, see Price (1989), Schueler (1991), Sobel and Copp (2001), Milliken (2008), and especially Frost (2014). Smith (1994, 209n8) himself goes on to recommend giving up talk of direction of fit, and instead "speaking directly about patterns of dispositions."

that the intention has the same two directions of fit. Or we can say instead that the normative belief has belief's direction of fit, and a rightful influence on our actions—and that the intention has intention's direction of fit, and is rightly influenced by our evaluations. Nothing seems to rule out the latter option. But if the latter option is right, it seems that belief and intention still essentially have different directions of fit. We then seem back to square one. How can an intention still itself be a belief?

Though it might seem counterintuitive, I think that the second of these options would be fine. In other words, the Identity View does not need to deny that intention and belief have opposite directions of fit, *with respect to their contents*. To see why, we can start with a simpler example.

Consider disbelief. A disbelief that *p*, as many philosophers have used the term, is a belief that not-*p*.[137] To deny that I went to the store today is to assert that I did not go; to disbelieve that I went is to believe that I did not go. Disbelief is clearly a species of belief. It is belief in a negation.

What is disbelief's direction of fit? Suppose I disbelieve that *p*, where *p* is "I returned my book today." But now I am presented with reliable video footage proving that I was home all day, and never went out, to the library or otherwise. I now recognize conclusive evidence against *p*. Do I give up my disbelief that *p*? Should I? Clearly not.

In one sense, disbelief can be said to aim to match the world. But it does not have belief's direction of fit, as direction of fit is often described. We can, of course, say that disbelief in *p* has belief's direction of fit, with respect to *not-p*. But the import of this fact is not immediately obvious. After all, not-*p* is not what I disbelieve. What I disbelieve is *p*. And the causal and normative relations we are interested in when we talk about direction of fit are not easily found between disbelief and what is disbelieved.

This confusion dissipates somewhat when we switch to thinking about negative beliefs. A negative belief is a belief that not-*p*, for some *p*. Negative beliefs clearly have belief's direction of fit with respect what is believed: not-*p*. And they also stand in some relations to a logical component of their content: *p*. Roughly speaking, negative beliefs tend to be, and ought to be, given up when *p* is seen to be true.

How might we compare the direction of fit of negative belief with that of disbelief? We might say that both of them have belief's usual direction of fit with respect to both *p* and not-*p*, since it is always the mind that adapts to accord with the world. But this way of speaking is optional. Again, consider Smith's dispositional conception of direction of fit. Does disbelief in *p* tend to go out of existence upon the perception that not-*p*? Clearly not; quite the contrary. Nor is it appropriate to give up one's disbelief upon finding that *p* is false.

Does this show that disbelief is an importantly different kind of state, distinct from belief? To disbelieve that *p* is not to believe that *p*, but to believe that not-*p*; and belief and disbelief react quite differently to, for example, evidence that *p*. So how can they possibly be the same state? The problem is spurious. Disbelief is belief-not. To be a disbelief is to

[137] In ordinary speech, "disbelief" can also refer to a refusal or inability to believe, rather than belief in a negation.

be a belief of a particular kind; and the 'content' of the disbelief is a component of the content of the belief. The state has different relations to *p* and to not-*p*; but however we describe those differences, calling belief-that-not-*p* "disbelief that *p*" does not lead to the troubling conclusion that the two cannot be identical. It merely shows that a single state, considered under different aspects, must sometimes be described in different ways.

The relation between intention and normative belief can be understood in an analogous way. On some ways of understanding direction of fit, intention and normative belief cannot be said to each have two different directions of fit, or to each have the same direction of fit as the other. But this is no obstacle to understanding the relation between them by analogy to the relation between disbelief and negative belief.

To disbelieve is to believe-not; on the Identity View, to intend is to believe-that-I-ought. To disbelieve that I went to the store today is to believe that I did not go; to intend to go is to believe that I ought to go. To put it more generally: to intend to *x* is to believe that *p*, where *p* is "I ought to *x*." If the Identity View is right, the intention and the belief are a single state, considered under different aspects. Unsurprisingly, the state has different relations to *x* and to *p*. I think little hangs on whether all of these relations, or some, or none of them, are described in terms of direction of fit. The basic facts remain the same: for the most part, we aim to believe what is true, to disbelieve what is false, to do what we should, and to make the world as we intend it to be. Considerations about direction of fit present no significant obstacle to thinking of disbelief as a species of belief. For the same reasons, they present no significant obstacle to thinking of intention as a species of belief. Considering the intention to *x*, and the belief that I ought to *x*, we have seen no fundamental differences in their relations, either to my *x*-ing, or to the facts about whether I ought to *x*.

Instead, I think we have seen an important feature of the Identity View spelled out in more detail. On the Identity View, intention is what we might call a *partially content-specifying state*. Like disbelief, an intention is a belief with a particular kind of content. Disbeliefs are beliefs that not *p*, for some *p*. Any belief whose content cannot be formulated as "not *p*" cannot be a disbelief. Like denial in the case of assertion, the state of disbelief is a state of belief with a particular kind of content. Similarly, on the Identity View, intentions are beliefs that I ought to *x*, for some action *x*. Any belief whose content cannot be formulated as "I ought to *x*," where *x* is an action, could not be an intention. The state of intention is a state of belief with a particular kind of content. If talk of direction of fit is confusing, it is especially so in the case of states like intention and disbelief, since, in each case, the 'content' of the partially content-specifying state is different from the 'content' of the genus state. In other words, what is disbelieved or intended is not what is believed, even if the disbelief or intention is itself a belief. We must therefore be careful to distinguish the 'content' of the intention, considered as an intention, from its 'content' when considered as a belief. The need for this distinction is not a sign of incoherence or any other theoretical flaw, since the distinction is one we already have to make in the case

of disbelief. Instead, it points to a structural characteristic of intention, according to the Identity View. Intentions, we can say, are beliefs-that-I-ought.[138]

III. Voluntary Control

In some ways, we do have voluntary control over our beliefs. If I want to believe that a dark room is brightly lit, all I have to do is flip the light switch. If I want to give up the belief that two plus two is four, suicide is an extreme but effective means. But many people have thought that, in an important sense, it is impossible to believe at will—that is, impossible to *decide* to believe, and to believe by deciding.[139] At most, it seems, one can *cause* oneself to believe—or not believe—by bringing about a change in oneself or in one's surroundings. We can flip a light switch, pull a trigger, or, more interestingly, attend to the evidence for a belief we do not yet have. And we can have beliefs that are influenced by our desires, as many of us do in the wishful belief that we are better than average drivers. But it is at least hard to imagine how we could intentionally form a belief, the way we can intentionally raise an arm or flip a light switch. It does not seem open to us to come to believe something intentionally, except by a kind of indirect causal influence.

Our lack of voluntary control over our beliefs is controversial, and difficult to describe with precision. But it is an important source of conviction that belief and intention are different kinds of state. It is natural to think: I cannot just decide to believe that I returned my book today, or that two plus two is five, or that the wall in front of me is red. And this inability seems to bring out a way in which belief is not a 'practical' state. I *can* decide to return my book today, and act on this decision. But I have no such control over my belief that I did. Normally, if I want to believe I returned my book, I will have to actually do it. And once I do, I will have no more of a choice about whether to believe I did than I had to begin with.

---

[138] Two additional points that I go on to make in §IV might be useful to anticipate here. First, I do not claim that the *concept* of intention is a partially content-specifying belief concept. Second, the status of intention as a partially content-specifying belief state offers a way to address concerns about whether intentions have propositional objects—that is, whether intentions are intentions that *p*. On the Identity View, all intentions have propositional objects, when considered as beliefs. All intentions are beliefs-that. But there is no tension between this thought and the observation that intentions are normally intentions "to" rather than "that". Intentions are intentions-to, because they are beliefs-that-I-ought-to.

More difficult issues arise in the case of talk of intentions-that, like my intention that my child go to college. I leave these aside here, though one point should by now be obvious. If such intention descriptions are elliptical ways of referring to an intention *to*, say, ensure that my child goes to college, the intention-to will nevertheless be propositional, when considered as a belief that I ought to ensure that my child goes to college. We can thus accommodate both the propositional object, and the sense of strangeness of talk of intending-that.

[139] See especially Williams (1970), Winters (1979), Hieronymi (2006), and Setiya (2008). For defenses of the possibility of believing at will, see Montmarquet (1986) and Ginet (2001).

These considerations about belief suggest an objection to the Identity View, which we can call *the voluntary control objection*. The objection is that intention cannot be a species of belief, because we lack voluntary control over our beliefs, in a way we do not over our intentions. As might by now be expected, I think our capacity for voluntary control over our intentions can be understood in the same way as in the case of normative belief. In this section, I will try to say why.

The examples I gave in motivating the voluntary control objection were misleading, in two ways. First, they focused on non-normative beliefs, like the belief that I returned a book, or that two plus two is four, or that the room I am in is brightly lit, or that the wall in front of me is red. These are mostly irrelevant. The topic is our lack of voluntary control over our normative beliefs, like the belief that I should return the book. Second, the examples compared beliefs not with intentions, but with actions, like raising an arm, flipping a switch, pulling a trigger, or returning a book. These actions seem clearly under our voluntary control; indeed, they are the kind of examples we give when we want to explain what "voluntary control" means. But the supposed disanalogy is with intention, not with action. The question to consider is: do we lack voluntary control over our normative beliefs in a way we do not lack voluntary control over our intentions? When we consider intention, I think we will find that our voluntary control is much more limited that it is over our actions.

Consider a now classic puzzle, known as *the toxin puzzle* (Kavka 1983, 33-34):

> You have just been approached by an eccentric billionaire who has offered you the following deal. He places before you a vial of toxin that, if you drink it, will make you painfully ill for a day, but will not threaten your life or have any lasting effects. (Your spouse, a crack biochemist, confirms the properties of the toxin.) The billionaire will pay you one million dollars tomorrow morning if, at midnight tonight, you intend to drink the toxin tomorrow afternoon. He emphasizes that you need not drink the toxin to receive the money; in fact, the money will already be in your bank account hours before the time for drinking it arrives, if you succeed.[140]

As Kavka emphasizes, the deal is appealing, and can seem easy. All you have to do is have an intention at midnight, and you get a million dollars. You would (he later stipulates) be willing to actually drink the toxin, but you do not even have to do that. But therein lies the puzzle, and, for you, the financial problem. If you get the million dollars at all, you will have it in your account well before you would need to actually drink the toxin. And you make no commitment to drink it—only to intend to. So as Kavka (1983, 34) puts it, "You had been thinking that you could avoid drinking the toxin and just pocket the million. But you realize that if you are thinking in those terms when midnight rolls around, you will not be intending to drink the toxin tomorrow."

[140] The toxin puzzle is a well-known and influential puzzle. For discussion, see Mele (1992b; 1995), Bratman (1999c), Andreou (2004), Clarke (2007), Gauthier (2008), Shah (2008), Levy (2009), and Tenenbaum (2009, 108-116).

As it turns out, it is not so easy to intend to drink the toxin. If you could, you would gladly form the intention. Most of us would be glad to intend it *and* drink the toxin. (If you would not be, you can upgrade the deal to a billion dollars, either to you or to your favorite charity.) The catch is that you know that drinking the toxin does nothing but cause you a day of needless pain. There will be no point to drinking it; you will have no reason to do it, and you know this now. For Kavka (1983, 35), this is why, much as you would like to form the intention to drink the toxin, "you cannot do so (or have extreme difficulty doing so)." As Kavka points out, "if intentions were simply…volitions fully under the agent's control, there would be no problem"(35). In fact the problem is daunting, despite the overwhelming utility of the elusive intention.[141]

Kavka's conclusion might seem too quick. Might there not be a way to form the intention to drink the toxin? He himself suggests there might be "extreme difficulty" rather than impossibility. Could it not be even a little easier than that? Perhaps we could ignore the pointlessness of following through on the intention, or form a determination to follow through "to make an honest proposition out of it" when the time comes. One might doubt that the obstacles to having the intention at midnight are so extreme.

For our purposes, there is no need to insist that such an attempt to form the intention could succeed only in extreme circumstances. Once again, the aim is to show that intention and normative belief can be given a parallel treatment. The relevant point here is that the same kind of doubt can be raised about our lack of voluntary control over our normative beliefs.

There is, more broadly, a range of doubts that can be raised about the impossibility of voluntary control over our beliefs generally. If I believe that no one likes me, can I not decide, with our without evidence, to believe that people do like me? A friend of the akratic anorexic believer in Chapters 2-4 might urge him: you have to decide to believe you're thin; hold on to that belief, no matter how it feels; it's important; you have to. Like intention, belief generally can be thought to be more directly amenable to decision, at least in some cases.

In the case of normative belief, we can consider the belief that I ought to drink the toxin. Once I have that belief, actually drinking the toxin is no longer impossible or especially difficult—and neither is having the intention to drink it, whatever we think about the intention's relation to the belief. I might thus try to convince myself that I ought to drink it; I might tell myself, for example, that I should have the intention and drink the toxin, too, as part of a "single package" which I can only choose as a package, and which is overwhelmingly worth it. The reasoning is questionable, but, I might say in the moment,

[141] Kavka (1983, 35) concludes that "intentions are better viewed as dispositions to act which are based on *reasons to act*." Though I say little about reasons, Kavka's conclusion is not far from mine. And the only work by another philosopher mentioned in the text of Kavka (1983) is Davidson (1980b), which, as we saw in Chapter 1, is a classic example of an Identity View. Kavka cites Davidson's paper as "an account that is generally congenial to the views presented here"(35).

never mind that.[142] This would be a way of forming the belief by deliberation, and also a way of forming the intention. On the other hand, I might come to the conclusion that I will just have to decide to believe I ought to drink it, and refuse to reconsider. (Such a brute decision might be the only way I see of getting the money.) Doing this might be difficult. But one might wonder whether the difficulty is really so extreme.

Voluntary control seems impossible, or at least difficult, over our intentions and over our normative beliefs. One might doubt the difficulty in both cases. Once again, we see a strikingly parallel set of issues. It can be hard to see how we can simply decide to intend to, or decide to believe we ought to, do something. And the subtleties of belief and intention formation might be argued to allow some leeway in either case. My conclusion is modest. The controversy has not been settled in either case, and I have not shown conclusively that a difference cannot be found between voluntary control over intention and normative belief. But we have seen that being a 'practical' state does not require being formed 'at will', the way a light switch can be flipped at will. Intentions are themselves responsive to our evaluations of a situation, and those evaluations cannot be changed at will—not, at least, unless our beliefs can too. Once again, we have found no clear difference between intention and normative belief with respect to voluntary control. Someone might come up with an original way of showing us something important about our voluntary control over intention or belief. But when she does, we must still ask whether the new insight shows a disanalogy. So far, we have still seen no reason to treat intention as different from normative belief.

IV. Reasoning

If intentions are not beliefs, it can seem puzzling that intention is integrated into a broader range of cognitive activity. Suppose you do believe you ought to return a book to the library. How do you come to intend to bring the book back? Do you form the intention on the basis of the belief? If so, how? By hypothesis, the intention is not itself a belief; so it cannot simply be the result of an inference from the belief, the way a further belief might be. But then the connection between the belief and the intention can seem obscure. How is intention integrated into our reasoning and cognition more generally? If practical reasoning, or reasoning about what to do, is a not a kind of reasoning through belief, what is it?

There are, of course, ways to address these questions. Perhaps intention is a kind of desire, or a state distinct from both belief and desire, distinguishable by its unique function or in some other way. Though I doubt that alternative theories can adequately characterize the cognitive integration of intention into the rest of our activity, I will not argue against competing views here. I want instead to highlight the especially clear and direct answer that the Identity View can offer. One of the advantages of thinking of intention as

[142] For the reasoning, see Gauthier (2008). For an argument that the reasoning is questionable, see Shah (2008, 16n36). For the view that the "never mind that" might be rational, see Andreou (2004).

normative belief is the straightforward picture this view can give of the cognitive integration of intention.

The straightforward picture is this: intention is itself the conclusion of reasoning about what one should do. If you start with a belief that you should return the book at some point, and come to see that today is the only day you can do it, you will likely conclude that you should return it today. On the Identity View, the belief you reach through this reasoning, that you should return the book to the library today, is your intention to return the book to the library today. No further inference or causal process is required, and no distinction between 'theoretical' and 'practical' reasoning is called for. I think this is an especially straightforward picture—both of how intention is integrated into reasoning, and of how that reasoning is 'practical'. But of course, straightforwardness would be a disadvantage, if this straightforward picture failed to accommodate an important characteristic of either intention or reasoning. It might help to look at one apparent example of a complexity that this view misses.[143]

In "How Action Governs Intention," Nishi Shah advances a "hypothesis" about "practical deliberation," or "deliberation that concludes in an intention" (Shah 2008, 1-2). Shah begins by observing that such deliberation typically focuses on what to do, often in the future, rather than on what to intend now. We saw this in considering the toxin puzzle: we normally form intentions in response to considerations about how we ought to act, not about what we ought to or would like to intend. On the other hand, other considerations can still have a causal influence on our intentions, without having an acknowledged role in deliberation. Like wishful thinking in the case of belief, for example, our desires can affect what we intend in ways we do not even notice. Why can't these other considerations be acknowledged in deliberation? Why are we unable to form an intention to drink the toxin on the basis of the desire to have that intention, even though that desire might influence our

[143] Another example can be found in the work of Pamela Hieronymi. In a series of papers, Pamela Hieronymi has argued that states like belief and intention can be understood as embodying answers to questions, and the reasons for each can be distinguished by the questions they bear on. One might then think that intentions embody answers to questions about *what to do*, while the corresponding normative beliefs embody answers to questions about *what we ought to do*. Intentions and normative beliefs can then be distinguished as answering different questions. All of Hieronymi's work is connected in one way or another to these issues, but see especially Hieronymi (2005 and 2006).

Though I have not argued in these terms in the text, I am in effect arguing that there is no viable way to show that the relevant questions are distinct. Once we answer the question of what to do, we might not do it; but this does not mean there is a further intelligible question of what to do. As Gibbard (2003, ix-x) puts it, "Thinking what I ought to do is thinking what to do." In this I follow Gibbard, though I do not take a stand in this chapter on many central issues with which he is concerned.

If we intentionally do something other than what we believe we ought to do, does our action or intention not embody an answer to some other question? Here my answer will not be surprising. Our action or intention can embody a *conflicting* answer to the *same* question. We can thus apply the arguments of Chapter 4, without giving up the idea that actions or intentions embody answers to questions.

deliberation in other ways? Shah's hypothesis is that "the concept of intention includes a standard of correctness"(2008, 2). According to this standard, an intention is "correct if and only if it is not the case that one ought not perform the action that is its object"(2008, 12). Shah argues that his hypothesis best explains why concluding deliberation whether to *intend* to do something requires answering the question whether to do it. Shah's hypothesis is also meant to explain why other considerations, such as the desirability of having a particular intention, cannot have an acknowledged role in practical deliberation, despite their ability to exert a causal influence on intention.

In defending his own view, Shah rejects the view, familiar from Davidson (1980b), that "intending to A is identical to judging that one ought to A"(Shah 2008, 14). He raises two objections to this view. First, he writes, "One can intend to A even though one thinks that one ought not A. This is what happens in cases of one type of akrasia"(11). I replied to this sort of objection at length in Chapter 4. The reply should by now be familiar: one can intend to A even though one thinks that one ought not A, if one *also* thinks one ought to A. Shah's first objection depends on ruling out the possibility of conflicting beliefs, judgments, or thoughts. Having considered the counterexample of *akrasia* in detail earlier, we can leave it aside here.

"Second," Shah (2008, 11) writes:

> one can believe that *p* or intend to *A* without having settled any question at all. Trying to settle a question (*i.e.*, deliberation) is an activity that we engage in sometimes when we form beliefs and intentions, but certainly not always….Furthermore, we attribute beliefs and intentions to others in order to explain their behavior or other mental states of theirs, without implying…that these states constitute answers that subjects have arrived at to questions that they have attempted to settle. As theorists we thus should not introduce such an endorsement or awareness into our account of the nature of these states.

It is not obvious how to understand Shah's notion of "settling" a question. Shah himself makes matters more complicated in this passage, by shifting from talk of settling to talking of 'trying' to settle. But the basic idea of his second objection is not hard to make out. Intentions are not always arrived at deliberatively, or explicitly or self-awarely endorsed. If we think of judgment as involving self-aware, explicit, or deliberative endorsement, it follows that intentions are not themselves judgments.

Shah's objection raises a complex set of issues, but I think it can be answered relatively simply. Like the notion of "settling" a question, the notion of judgment can be difficult to understand without more detailed investigation. But we do not need to undertake that detailed investigation here. Instead, we can consider a prior question. Why consider *judgment* at all? As Shah's own objection suggests, belief is the more natural analogue to intention. His own parallel treatment of belief and intention stands in explicit contrast to judgment. As he puts it (13n29): "Beliefs and intentions are mental attitudes, whereas judgments are mental acts." If this contrast is right—and if, as it assumes,

attitudes are not acts—then this fact alone is enough to show that intentions are not judgments. But then the natural alternative to Shah's hypothesis from the start is that intentions are identical, not with judgments, but with beliefs. I formulated the Identity View in terms of identity with belief, partly in response to considerations like these. And as we have seen, neither of Shah's objections is compelling against the view that intentions are normative beliefs.

Shah's own hypothesis remains attractive. A standard of correctness included in the concept of intention might indeed help explain the centrality of normative considerations in practical deliberation. But this hypothesis is compatible with the Identity View. Indeed, the Identity View can itself offer an explanation for the standard of correctness. If an intention to *A* is a belief that one ought to *A*, it is natural to think that the intention is correct if and only if one actually ought to *A*. It is thus open to the Identity View, and to Shah, to look for a happy synthesis, in which the Identity View supports and explains the standard of correctness.

Indeed, on this combined view, the Identity View can offer an improved formulation of the standard itself. On Shah's formulation, an intention is "correct if and only if it is not the case that one ought not perform the action that is its object"(12). This formulation is permissive in form. As Shah puts it (12n27):

> This formulation of the standard of correctness allows decisions in Buridan's Ass cases. Although suspension of belief is rationally required when one's evidence equally supports two opposing hypotheses, suspension of action is not required when one is faced with two equally desirable options; one is rationally permitted to pick either one. The standard of correctness is met by either act because it is not the case that either act is such that one ought not to perform it.

Shah thinks that Buridan cases, in which we are faced with equally desirable options, force a weaker formulation of the standard of correctness. But as we saw in Chapter 5, they do not. Our resolutions of Buridan cases are compatible with the Identity View. We can thus have a simpler view—intentions are normative beliefs—and a simpler standard of correctness: an intention to *A* is correct and if only if one ought to A. The Identity View can thus be combined with Shah's hypothesis, and help in both defending and formulating the hypothesis itself. Shah thus offers no significant threat to the Identity View. His objections to it can be answered, and his own hypothesis is compatible with it.[144]

---

[144] Shah's hypothesis parallels his view that the concept of belief includes a standard of correctness: belief is correct if and if only if it is true. See Shah (2003), Shah and Velleman (2005), and Shah (2008, 12). What I am describing is a more closely parallel standard of correctness for intention: intention is correct if and only if one ought to perform the action that is its object. That is, my intention to *x* is correct if and only if I ought to *x*. The Identity View can explain this standard as an instance of belief's own standard. My intention to *x* is correct if and only if I ought to *x*, because my intention is a belief that I ought to *x*, and the belief is correct if and only if it is true.

One might also ask, in this context, what deliberation among permissible options could be. How do I choose a breakfast cereal, if five different cereals are permissible, and the standard of

Still, in defending the Identity View, I do not insist that Shah's hypothesis is correct. On the contrary, I suspect that it is not the right way to understand the Identity View, or the centrality of normative considerations in deliberation. For Shah (2008, 15), "The conception of deliberation that emerges is this: deliberation is reasoning aimed at issuing in some result in accordance with norms for results of that kind." Three concerns might be raised here. First, as Shah has emphasized, deliberation can focus on what to do, without using the concept of intention. The standard of correctness might then be better explained as a standard involved in *deliberation* toward intention, rather than as a standard included in the concept of intention. Second, deliberation may not consist entirely in reasoning at all. What we call 'deliberation' can include processes such as 'mulling over' an idea, by attending to it without much reasoning at all, or even staring into space and hoping clarity dawns on us. A standard of correctness included in the concept of intention would have even less to say here about why some considerations have a causal influence on our intentions without being able to play an acknowledged role in deliberation. Third, it is not clear that concepts like 'intention' and 'belief' do include a standard of correctness. Those who think of such states in behaviorist or functional terms might vehemently deny that they include any such standard. If they are right, then Shah's hypothesis would be explanatory if true, but it would not be true.

I think the Identity View is best understood as putting forward a metaphysical truth, rather than a conceptual one. The concept of intention might be too vague, or understood too much in functional terms, to carry with it much more than the attribution of an aim to a person. But we can still ask what the attribution of an aim *to a person* amounts to. According to the Identity View, the key difference between an intention and the 'aim' of a plant or liver is that an intention involves—indeed, is—normative awareness: a belief that one ought to perform an action. This view offers a substantive account of what intention is. And I think it offers a compelling picture of practical reasoning. Practical reasoning is reasoning toward intention, which *is* reasoning toward a conclusion about what one ought to do. If this view is right, there is no need to search for an alternative kind of reasoning that practical reasoning could be. And once again, there is no compelling disanalogy

---

correctness offers no guidance? On the Identity View, we have an answer to this question. The standard does offer guidance. The permissibility in question is not 'moral' in some narrow sense of 'moral'; if we think we should just eat whichever one we feel like eating, this 'should' offers a standard in this case. And as we saw in Chapter 5, if we see no distinguishing qualities at all, we are naturally led to conclude that we should act nonintentionally to determine which one we will eat. This solution to Buridan cases is precisely a way of deliberating between permissible options, when the options are 'permissible' in the relevantly narrow sense. On Shah's stated view, on the other hand, an intention is "correct if and only if it is not the case that one ought not perform the action that is its object"(2008, 12). Presumably, an intention to eat any of the five cereals would be correct. But then how does the norm govern deliberation? Avoiding the conclusion that an intention would be incorrect does not tell us how deliberation can proceed. I think the solution I offer in Chapter 5 can be adapted to Shah's view. But if this is right, then, as I say in text, the motivation for weakening the standard to be permissive disappears. There is no obstacle to treating intention's standard of correctness as more closely parallel to belief's standard.

between intention and belief, except where normative belief itself differs from some of our other ordinary beliefs.

It might seem that a more fundamental disanalogy between theoretical and practical reasoning is that theoretical reasoning reaches a conclusion that can be true or false. Can intentions be said to be have a truth-value? I have not argued that normative beliefs themselves are states with truth-values, rather than some non-cognitive state unlike our ordinary beliefs. But I am inclined to resist a non-cognitive view, and anyway the Identity View should ideally not rely on one. So it is worth asking, at least briefly, whether an intention can be false or true.

I believe that it can, and that this fact presents no special problem for the Identity View. Defending this view fully is a larger project.[145] But I think something can be said about why the idea of the truth or falsity of an intention might seem strange.

As an analogy, consider non-religious belief-in. Parents tell their children: "I believe in you." I believe in my husband, my students, and my president. How are these 'beliefs' related to ordinary beliefs, which are always beliefs 'in' a proposition? Are all beliefs-in, at bottom, a kind of belief-that? Is believing in someone the same as believing that she is likely to succeed, or a good person, or worthy of support? These questions are not easy to answer; I, for one, am unsure what to say about them.[146] But now consider a different question. Can a parent's belief in her child be true or false? It seems strange to think that it can. But here is the important point: talk of truth or falsity seems strange here, *whether or not* belief-in is in fact belief-that. If a belief in one's child can indeed be shown to be a belief-that, it will indeed be true or false. And still this capacity for truth or falsity should strike us as strange.

This sense of strangeness has several sources, which I think can also be seen in the case of intention.[147] One is relatively superficial: talk of truth or falsity is unusual and often inappropriate in the context at hand. When we talk of believing in someone, we might be encouraging them, or simply deriving happiness from the fact of our belief in them. Similarly, when we talk about our intentions, we often focus on their role in action and

---

[145] For example, one might object that evaluation of beliefs as true or false is binary, while evaluation of intentions as required, permitted, or forbidden is ternary. This objection is doubly difficult to make compelling. First, evaluation of "This man is bald" or "This sentence is false" may well yield a third result that is neither truth nor falsity. Second, a growing number of "epistemically permissivist" philosophers have thought that many beliefs are permitted, rather than required or forbidden. It is thus not clear, either that evaluation for truth is binary, or that beliefs are evaluated for truth as opposed to for requiredness or permissibility. I am optimistic that belief and intention can be given a parallel treatment here, but that treatment will have to be complex, and I leave the details for another occasion. Other views on which intentions are beliefs and so, presumably, have truth-values include Velleman (1989), Setiya (2007), and McDowell (2010), though there has been little discussion of the apparent oddity of thinking of intentions as true or false.

[146] For some of the controversy, see Price (1965), MacIntosh (1970, 1994), and Williams (1992).

[147] Much of what I say can also be applied by those who want to make plausible that knowing-how is a kind of knowing-that; but I leave this aside here.

coordination. Thoughts of truth or falsity can thus seem irrelevant, and therefore strange or out of place.

Two further sources of apparent strangeness lie in the object and the grammar of each state. When we believe in a person, the object of our belief cannot be true or false. When we intend to return a book, the action we intend to perform cannot be true or false.[148] People and actions do not have truth-values. Relatedly, we say we believe 'in' someone and intend 'to' return a book, rather than believing or intending 'that'. The preposition seems to mark the fact that we are dealing with a state that is not propositional, and therefore cannot be true or false. If belief-in or intending-to are in fact beliefs-that, the underlying structure of a propositional state must be reconciled with an apparently quite different grammar and object. In the case of intention, I have offered such a reconciliation. Belief-that-I-ought is unproblematically followed by 'to' and an action, and is unproblematically a form of belief-that.

I am inclined to think that belief-in is not, at bottom, a kind of belief-that. This is in part because, in the case of belief-in, it is hard to know *which* proposition someone with a belief-in would be believing. I do not know what belief-that would fit the phenomena of believing in someone. In the case of intention, I have argued that there is a single belief-that which can be identified with the intention. In the other ways I mentioned, it still naturally seems strange that an intention can be true, the way beliefs can. But this strangeness can be taken as a sign of an interesting and surprising result, rather than an impossibility. We have seen no reason to think that intentions could not in principle be true or false, though we have seen why it might be awkward to speak directly of their truth or falsity. Instead, I think the identity of intention and normative belief about action has emerged as a simple, striking, unobvious hypothesis. It is a hypothesis with far-reaching implications, and, I tentatively conclude, with no compelling objections to it.

V. Conclusion

The Identity View offers a conception of what intention is. Intention is a first-personal, normative belief about one's own action: a belief that I should, or ought to, go to the store, return a book, and so on. It has wide-ranging and, I think, defensible implications about topics such as practical reasoning, voluntary control, direction of fit, *akrasia*, Buridan cases, and *accidie*. I have only begun to develop the Identity View into a general theory of intention. But if we return to the motivations for the guise-of-the-good views considered in Chapter 1, §I, we can see how well the Identity View does justice to these motivations.

First, the Identity View offers a conception of what is distinctive about intention and intentional action. The sense that evaluation plays a central role in our intentional activity is given straightforward expression by the thought than an intention is itself a

[148] People and actions *can* be true or false, in the sense of loyal or disloyal; but I leave this aside.

normative belief. We can see straightforwardly why acting on an intention is a form of self-government: we govern ourselves by doing what we believe we ought to do, rather than what our inclinations might tempt us toward. And the fact that someone believes she *ought* to do something offers a way of explaining the attribution of the intention to this person. The intention is hers, because it behaves as her normative beliefs do; it is responsive to reasons, felt with conviction, reasoned from in a range of circumstances, reported to others and used in interpersonal coordination, and so on. The Identity View tells us what makes something an intention, in a way that does justice to central features of intention.

Second, the Identity View offers a unified conception of the explanation of particular actions. As we saw in Chapter 1, it is odd to ask questions like: "I see why she thinks she should buy these clothes, but why is she buying them?" The explanation of why she thinks she should buy clothes is itself an explanation of why she buys them. The Identity View does justice to this fact. On the Identity View, the explanation of why someone believes she should buy clothes is itself the explanation of why she buys them, because her belief is itself her intention. There is no further need to explain her intention; and except in unusual circumstances, there is no further need to explain why someone does what she intends to.

Third, the Identity View does justice to our sense of a parallel between belief and intention or intentional action. Both, it seems, are in some sense trying to get things right: belief with respect to the true, and intention and intentional action with respect to what we ought to do. On the Identity View, this parallel is a straightforward consequence of the nature of intention. If intentions are beliefs, then of course intention 'tries' to get things right, in whatever sense belief does. And there is no disanalogy between the 'true' and the 'good' or 'ought' here. Beliefs aim to get things right about their subject matter, which, in the case of intentions, is what we ought to do.

Fourth, the Identity View does justice to the historical precedent of earlier guise-of-the-good views. If the Identity View is right, then earlier adherents of guise-of-the-good views were indeed on to something important. Indeed, if the Identity View is right, these earlier adherents tended to hold weakened versions of the correct view. They recognized a fundamental connection between intention or action and belief, but were convinced by *akrasia*, Buridan cases, direction of fit, an emphasis on self-conscious judgment, and other examples and theoretical objections that seemed to tell against ambitious guise-of-the-good views. By responding to these examples and objections, the Identity View can build on historical precedent, while improving our understanding of the examples and theoretical issues.

Fifth, the Identity View has the potential to play a central role in foundational arguments in moral philosophy. If our beliefs about what we ought to do are themselves intentions, then a conclusion we reach about what we ought to do will have an inescapable hold on us. If we can establish that some principle or value is one that we should act on, acting as we should will not simply be optional, in the sense of being an alternative we may or may not intend to take. Foundational arguments in moral philosophy may be able

to appeal to the Identity View in addressing themselves to us. I have not tried to show that such arguments can be successful; but the potential to play a role in such arguments is another motivation for being interested in guise-of-the-good views, and the Identity View has this potential.

Sixth, the Identity View can, like guise-of-the-good views generally, be seen as an extension of ordinary charity or generosity of interpretation. It says that even when we are conflicted, confused, or exhausted, we intend to do what we believe we should. We are all, in this sense, well-intentioned. One can still say that the road to hell is paved with good intentions; after all, on the Identity View, *all* of the most heinous actions, when intended, were ones their doers believed they ought to perform. But they are at least not done without regard for what ought to be done. They embody terrible mistakes in normative thinking, rather than a lack of interest in its conclusions.

I began this dissertation by introducing these motivations for holding or being interested in guise-of-the-good views. As I have emphasized, the Identity View can be thought to 'overdo' justice to these motivations, offering a theory that is too unified and too 'generous'. I have tried to articulate, distinguish, and answer these doubts. If the preceding chapters are right, the Identity View does not in fact overstate the parallels between intention and belief, or miss important divergences between evaluation and motivation. Most of these chapters have been devoted to arguing that the Identity View does not go too far in doing justice to these motivations for guise-of-the-good views. But if it does not go too far, I think it does go far enough. The Identity View is deeply motivated; and it does justice to a wide range of motivations for it, while also shedding light on the details of the examples and theoretical concerns that were supposed to tell against it.

The Identity View cannot possibly be thought to be not unified enough in its treatment of intention and belief. It claims no disanalogy at all between intention and belief; it claims identity. But the Identity View might still seem not 'generous' enough. It might seem distinctly ungenerous, uncharitable, or unkind to treat all intentions as normative beliefs. On the Identity View, whenever we intend to do something, we believe we ought to do it; and whenever we fail to do what we intend, we have failed to do what we believe we ought to do. The guilt and blame can seem crippling. Can such a moralized view be generous? Can it be true that every time we fall short of an intention, we open ourselves to the criticisms and hard feelings that come with failing to do what we ought?

Although the Identity View is an ambitious guise-of-the-good view, this objection brings out that, in one way, the Identity View claims very little. It includes no thought that we see what we intend as morally required, in any narrow sense of 'moral'. If I intend to set an alarm, I believe I should set it: not that I am bad if I do not, or that I owe it to someone to set the alarm, or that I ought 'morally', in some further sense of 'morally', to set the alarm. Nor does the Identity View make any claim about guilt, blame, or other forms of moral appraisal. Some might believe that guilt or blame are appropriate whenever someone fails to do what she believes she should. But it is this view, not the Identity View, that would be ungenerous or unkind. A less extreme view about the appropriateness of guilt and blame would presumably tell us which cases of failing to do what we intend

would make guilt or blame appropriate. Even a complete skeptic about the appropriateness of guilt or blame could accept the Identity View, and live a life infused with normative thought, without ever accepting these kinds of moral response as appropriate. The Identity View carries no implication that a person should be blamed or feel a certain way, in all cases of failure or even in one.

Guise-of-the-good views offer a way to understand what intentional action and intention are, in a way that does justice to the central role of evaluation. As a strategy for considering this large family of views, I began in Chapter 1 with the idea of taking an especially ambitious one, and seeing in what ways it needs to be weakened. I turned to the Identity View as an especially ambitious guise-of-the-good view; and I came to think, and have argued here, that it does not need to be weakened. It offers a conception of intention, action explanation, and the parallels between intention and belief that addresses a range of counterexamples and theoretical concerns, does justice to and builds on historical precedent, and offers a generous picture of human nature that promises to be relevant to foundational questions in moral philosophy. There is, I believe, no example in which intention and normative belief can be seen to come apart, and no general disanalogy to be drawn between them. My hope is that recognizing this fact will help us think about *what* we should be doing, without blaming each other and ourselves for our failures.

Bibliography

Where known, I include original dates of publication and translated titles in brackets.

Aamodt, Sandra, and Sam Wang (2008). "Tighten Your Belt, Strengthen Your Mind." *New York Times* Op-Ed, April 2, available online at http://www.nytimes.com/2008/04/02/opinion/02aamodt.html?_r=0. Last accessed on April 22, 2016.

Ackrill, J.L (1973). "Introduction." In Aristotle, *Aristotle's Ethics* (New York: Humanities Press).

Adler, Jonathan (2002a). "Akratic Believing?" *Philosophical Studies* 110:1, 1-27.

---------- (2002b). *Belief's Own Ethics*. Cambridge, MA: MIT Press, 2002.

Adler, Jonathan E., and Bradley Armour-Garb (2007). "Moore's Paradox and Transparency." In Green and Williams (2007), 146-162.

Al-Ghazali (1963 [1095]). *Tahafut Al-Falasifah* [*The Incoherence of the Philosophers*]. Tr. Sabih Ahmad Kamali. Lahore: Pakistan Philosophical Congress.

Albritton, Rogers (1995). "Comments on 'Moore's Paradox and Self-Knowledge'." *Philosophical Studies* 77:2/3, 229-39.

Andreou, Chrisoula (2004). "Instrumentally Rational Myopic Planning." *Philosophical Papers* 33:2, 133-45.

Anscombe, G.E.M (1957). *Intention*. Oxford: Blackwell.

Aristotle (1984). *The Complete Works of Aristotle*. Revised Oxford Translation, ed. Jonathan Barnes. Princeton: Princeton University Press.

---------- (1999). *The Nicomachean Ethics*. Tr. Terence Irwin. Indianapolis: Hackett.

Armstrong, David (1968). *A Materialist Theory of the Mind*. New York: Routledge & Kegan Paul.

Averroes (2008 [ca. 1180]). *Tahafut al-Tahafut* [*The Incoherence of the* Incoherence]. Tr. Simon van den Bergh. London: Gibb Memorial Trust.

Audi, Robert (1972). "The Concept of 'Believing'." *Personalist* 53, 43–62.

---------- (1979). "Weakness of Will and Practical Judgment." *Noûs* 13:2, 173-196.

Baldwin, Thomas (1990). *G.E. Moore*. New York: Routledge,

Baumeister, Roy F. (1989). "The Optimal Margin of Illusion." *Journal of Social and Clinical Psychology* 8:2, 176-189.

---------- (2002). "Ego Depletion and Self-Control Failure: An Energy Model of the Self's Executive Function." *Self and Identity* 1, 129-136.

---------- (2003). "Ego Depletion and Self-Regulation Failure: A Resource Model of Self-Control." *Alcoholism: Clinical and Experimental Research* 27:2, 1-4.

---------- (2012). "Self-Control: The Moral Muscle." *The Psychologist* 25:2, 112-115.

Baumeister, Roy F., Ellen Bratslavsky, Mark Muraven, and Dianne M. Tice (1998). "Ego Depletion: Is the Active Self a Limited Resource?" *Journal of Personality and Social Psychology* 74:5, 1252-1265.

Baumeister, Roy F. and Julie Juola Exline (1999). "Virtue, Personality, and Social Relations: Self-Control as the Moral Muscle." *Journal of Personality* 67:6, 1165-1194.

Baumeister, Roy F. and Todd F. Heatherton (1996). "Self-Regulation Failure: An Overview." *Psychological Inquiry* 7:1, 1-15.

Baumeister, Roy F, Erin A. Sparks, Tyler F. Stillman, and Kathleen D. Vohs (2008). "Free Will in Consumer Behavior: Self-Conrol, Ego Depletion, and Choice." *Journal of Consumer Psychology* 18, 4-13.

Baumeister, Roy F. and John Tierney (2011). *Willpower: Rediscovering the Greatest Human Strength*. New York: Penguin.

Baumeister, Roy F. and Kathleen D. Vohs (2007). "Self-Regulation, Ego Depletion, and Motivation." *Social and Personality Psychology Compass* 1, 1-14.

Baumeister, Roy F, Kathleen D. Vohs, and Dianne M. Tice (2007). "The Strength Model of Self-Control." *Current Directions in Psychological Science* 16:6, 351-355.

Björnsson, Gunnar, Caj Strandberg, Ragnar Francén Olinder, John Eriksson, and Fredrik Björklund, eds., *Motivational Internalism*. New York: Oxford University Press, 2015.

Bouyges, Maurice (1927). "Notice." In Algazel [Al-Ghazali], *Tahafot Al-Falasifat* (Beirut: Imprimerie Catholique).

Bovens, Luc (1995). "P and I Will Believe that not-P: Diachronic Constraints on Rational Belief." *Mind* 104:416, 737-760.

Bowden, Hannah (2012). "A Phenomenological Study of Anorexia Nervosa." *Philosophy, Psychiatry, & Psychology* 19:3, 227-241.

Boyle, Matthew (2009). "Two Kinds of Self-Knowledge." *Philosophy and Phenomenological Research* 78:1, 133-64.

Boyle, Matthew and Douglas Lavin (2010). "Goodness and Desire." In Tenenbaum (2010), 161-201.

Braithwaite, R.B. (1932). "The Nature of Believing." *Proceedings of the Aristotelian Society* 33, 129–146.

Bratman, Michael (1987). *Intentions, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

---------- (1999a). "Davidson's Theory of Intention." Reprinted in his *Faces of Intention: Selected Essays on Intention and Agency* (New York: Cambridge University Press), 209-24.

---------- (1999b). "Review of Korsgaard's *The Sources of Normativity*." Reprinted in his *Faces of Intention: Selected Essays on Intention and Agency* (New York: Cambridge University Press), 265-78.

---------- (1999c). "Toxin, Temptation, and the Stability of Intention." Reprinted in his *Faces of Intention: Selected Essays on Intention and Agency* (New York: Cambridge University Press), 58-92.
---------- (2007). "A Desire of One's Own." Reprinted in his *Structures of Agency: Essays* (New York: Oxford University Press), 137-161.
Brink, David (1989). *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.
Broadie, Sarah (1991). *Ethics with Aristotle*. New York: Oxford University Press.
Bromwich, Danielle (2008). *Belief Internalism*. Ph.D. dissertation, University of Toronto.
Burge, Tyler (1977). "Belief de re." *The Journal of Philosophy* 75, 119–138.
Byrne, Alex and Matthew Boyle (2011). "Self-Knowledge and Transparency." *Proccedings of the Aristotelian Society* Supplementary Volume 85, 201-41.
Callard, Agnes (2008). *An Incomparabilist Account of* Akrasia. Ph.D. dissertation, the University of California – Berkeley.
Cassam, Quassim (2010). "Judging, Believing, and Thinking." *Philosophical Issues* 20:1, 80-95.
Chang, Ruth, ed (1997). *Incommensurability, Incomparability, and Practical Reason*. Cambridge, MA: Harvard University Press.
Churchland, Paul M. (1981). "Eliminative Materialism and the Propositional Attitudes." *Journal of Philosophy* 78, 67–90.
Clarke, Randolph (2007). "Commanding Intentions and Prize-Winning Decisions." *Philosophical Studies* 133:3, 391-409.
Cohen, L. Jonathan (1992). *An Essay on Belief and Acceptance*. Oxford: Oxford University Press.
Dancy, Jonathan (1993). *Moral Reasons*. Cambridge, MA: Blackwell.
Davidson, Donald (1980a). "How is Weakness of the Will Possible?" Reprinted in Donald Davidson, *Essays on Actions and Events* (Oxford: Clarendon Press), 21-42.
---------- (1980b). "Intending." Reprinted in Donald Davidson, *Essays on Actions and Events* (Oxford: Clarendon Press), 83-102.
---------- (1982). "Paradoxes of Irrationality." In Richard Wollheim, ed., *Philosophical Essays on Freud* (Cambridge: Cambridge University Press), 289-305.
---------- (1985a). "Deception and Division," in E. LePore and B. McLaughlin, eds., *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (New York: Basil Blackwell), 138-148.
---------- (1985b). "Replies to Essays I-IX." In Bruce Vermazen and Merrill B. Hintikka, eds., *Essays on Davidson: Actions and Events* (New York: Oxford University Press, 1985), 195-229.
De Almeida, Claudio (2007). "Moorean Absurdity: An Epistemological Analysis." in Green and Williams (2007), 53-75.
Descartes, René (2008 [1641]). *Meditations on First Philosophy: with Selections from the Objections and Replies*. Tr. Michael Moriarty. Oxford: Oxford University Press.
Doris, John and Stephen Stich (2005). "As a Matter of Fact: Empirical Perspectives on

Ethics." In Frank Jackson and Michael Smith, eds., *The Oxford Handbook of Contemporary Philosophy* (New York: Oxford University Press), 114-152.

Dretske, Fred (1971). "Reasons, Knowledge, and Probability." *Philosophy of Science* 38:2, 216-220.

----------. (1988). *Explaining Behavior*. Cambridge, MA: MIT Press.

Dreyfus, Hubert, and Sean Dorrance Kelly (2011). *All Things Shining: Reading the Western Classics to Find Meaning in a Secular Age*. New York: Simon & Schuster.

Edgley, Roy (1969). *Reason in Theory and Practice*. London: Hutchinson.

Elga, Adam (2005). "On Overrating Oneself – and Knowing it." *Philosophical Studies* 123, 115-124.

Fodor, Jerry (1975). *The Language of Thought*. New York: Cromwell.

---------- (1981). *Representations*. Cambridge, MA: MIT Press.

Frankfurt, Harry (1988a). "Freedom of the Will and the Concept of a Person." In Harry Frankfurt, *The Importance of What We Care About* (New York: Cambridge University Press), 11-25.

---------- (1988b). "Identification and Externality." In Harry Frankfurt, *The Importance of What We Care About* (New York: Cambridge University Press), 58-68.

Gallois, André (2007). "Consciousness, Reasons, and Moore's Paradox." In Green and Williams (2007), 165-88.

Gauthier, David (2008). "Rethinking the Toxin Puzzle." In Jules L. Coleman and Christopher W. Morris, eds., *Rational Commitment and Social Justice: Essays for Gregory Kavka* (New York: Cambridge University Press, 1998), 47-58.

Gibbard, Allan (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, MA: Harvard University Press.

---------- (2003). *Thinking How to Live*. Cambridge, MA: Harvard University Press.

Ginet, Carl (2001). "Deciding to Believe." In Mathias Steup, ed., *Knowledge, Truth, and Duty* (New York: Oxford University Press), 63-76.

Giordano, Simona (2005). *Understanding Eating Disorders*. Oxford: Oxford University Press.

Goldstein, L (1988). "Wittgenstein's Late Views on Belief, Paradox, and Contradiction." *Philosophical Investigations* 11:1, 49-73.

Gombay, André (1988). "Some Paradoxes of Counterprivacy." *Philosophy* 63:244 (April 1988), 191-210.

Gosling, Justin (1990). *Weakness of the Will*. New York: Routledge.

---------- (1993). "Mad, Drunk, or Asleep? Aristotle's Akratic." *Phronesis* 38:1, 98-104.

Greco, Daniel (2012). "The Impossibility of Skepticism." *Philosophical Review* 121:3, 317-358.

---------- (2014). "A Puzzle about Epistemic Akrasia." *Philosophical Studies* 167:2, 201-219.

Green, Mitchell and John N. Williams, ed. (2007). *Moore's Paradox: New Essays on Belief, Rationality, and the First Person*. Oxford: Oxford University Press.

Hagger, Martin S., Chantelle Wood, and Chris Stiff (2010). "Ego Depletion and the Strength Model of Self-Control: A Meta-Analysis." *Psychological Bulletin* 136:4, 495-525.

Haidt, Jonathan and Fredrik Bjorklund (2008). "Social Intuitionists Answer Six Questions About Moral Psychology." In *Moral Psychology*, vol. 2, ed. Walter Sinnott-Armstrong. Cambridge, MA: MIT Press.

Haidt, Jonathan, Fredrik Bjorklund, and Scott Murphy (2000). "Moral Dumbfounding: When Intuition Finds No Reason." Manuscript online at http://commonsenseatheism.com/wp-content/uploads/2011/08/Haidt-Moral-Dumfounding-When-Intuition-Finds-No-Reason.pdf. Last accessed November 12, 2013.

---------- (2001). "The Emotional Dog and its Rational Tail." *Psychological Review* 108:4, 814-834.

---------- (2005). Interview in *The Believer*. August 2005. Available at http://www.believermag.com/issues/200508/?read=interview_haidt. Last accessed April 22, 2016.

Hájek, Alan (2007). "My Philosophical Position Says 'p' and I Don't Believe 'p'." In Green and Williams (2007), 217-31.

Halban, Emily (2009). *Perfect: Anorexia and Me*. London: Vermilion Press.

Hare, R.M. (1952). *The Language of Morals*. Oxford: Clarendon Press.

---------- (1963). *Freedom and Reason*. Oxford: Clarendon Press.

Heil, John (1984). "Doxastic Incontinence." *Mind* 93:360, 56-70.

Hieronymi, Pamela (2005). "The Wrong Kind of Reason." *The Journal of Philosophy* 102:9, 437-457.

---------- (2006). "Controlling Attitudes." *Pacific Philosophical Quarterly* 87, 45-74.

Hintikka, Jaakko (1962). *Knowledge and Belief*. Ithaca, NY: Cornell University Press.

Holton, Richard (2009). *Willing, Wanting, Waiting*. New York: Oxford University Press.

Horowitz, Sophie (2014). "Epistemic Akrasia." *Noûs*, 48:4, 718-744.

Huemer, Michael (2007). "Moore's Paradox and the Norm of Belief." In Susana Nuccetelli and Gary Seay, *Themes from G.E. Moore: New Essays in Epistemology and Ethics* (New York: Oxford University Press), 142-157.

Hume, David (2000 [1738]). *A Treatise of Human Nature*. New York: Oxford University Press.

Hurley, Susan (1989). *Natural Reasons*. Oxford: Oxford University Press.

Jacobson, Daniel (2012). "Moral Dumbfounding and Moral Stupefaction." In Mark Timmons, ed., *Oxford Studies in Normative Ethics*, vol. 2 (New York: Oxford University Press), 289-315.

Job, Veronika, Carol S. Dweck, and Gregory M. Walton (2010). "Ego Depletion—Is it All in Your Head? Implicit Theories about Willpower Affect Self-Regulation." *Psychological Science* 21:11, 1686-1693.

Kant, Immanuel (1997 [1788]). *Critique of Practical Reason*. Tr. Mary Gregor. Cambridge: Cambridge University Press.

---------- (1998 [1785]). *Groundwork of the Metaphysics of Morals*. Tr. Mary Gregor. Cambridge: Cambridge University Press.
Kaplan, David (1968). "Quantifying in." *Synthese* 19, 178–214.
Kavka, Gregory (1983). "The Toxin Puzzle." *Analysis* 43:1, 33-36.
Korsgaard, Christine (1996). *The Sources of Normativity*. Cambridge: Cambridge University Press.
---------- (1997). "The Normativity of Instrumental Reason." In Garrett Cullity and Berys Gaut, eds., *Ethics and Practical Reason* (Oxford: Clarendon Press), 215-254.
---------- (2008). "Self-Constitution in the Ethics of Plato and Kant." In her *The Constitution of Agency* (New York: Oxford University Press), 100-126.
---------- (2009). *Self-Constitution: Agency, Identity, Integrity*. Oxford: Oxford University Press.
Kriegel, Uriah (2004). "Moore's Paradox and the Structure of Conscious Belief." *Erkenntnis* 61:1, 99-121.
Leibniz, Gottfried Wilhelm (1948). *Textes Inédits*. Ed. Gaston Grus. Paris: Presses Universitaires de France.
---------- (1952 [1710]). *Theodicy*. Tr. E.M.Huggard. New Haven: Yale University Press.
Levy, Ken (2009). "On the Rationalist Solution to Gregory Kavka's Toxin Puzzle." *Pacific Philosophical Quarterly* 90, 267-89.
Levy, Neil (2004). "Epistemic Akrasia and the Subsumption of Evidence: A Reconsideration." *Croatian Journal of Philosophy* 4:10, 149-156.
Lewis, David (1972). "Psychophysical and Theoretical Identifications." *Australasian Journal of Philosophy* 50, 249–258.
---------- (1979). "Attitudes de dicto and de se." *Philosophical Review* 88, 513–543.
---------- (1980). "Mad Pain and Martian Pain." In N. Block, ed., *Readings in the Philosophy of Psychology, vol. 1* (Cambridge, MA: Harvard University Press), 216–222.
MacIntosh, J.J. (1970). "Belief-In." *Mind* 79, 395-407.
---------- (1994). "Belief-in Revisited: A Reply to Williams." *Religious Studies* 30:4, 487-503.
Malcolm, Norman (1958). *Ludwig Wittgenstein: A Memoir*. Oxford: Oxford University Press.
McAdam, James I. (1965). "Choosing Flippantly or Non-Rational Choice." *Analysis* 25: Supplement 3 (January 1965), 132-136.
McDowell, John (2010). "What is the Content of an Intention in Action?" *Ratio* 23, 415-32.
Mele, Alfred (1986). "Incontinent Believing." *The Philosophical Quarterly* 36:143, 212-222.
---------- (1987). *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control*. New York, Oxford University Press.
---------- (1991). "Motivational Ties." *Journal of Philosophical Research* 16, 431-442.

---------- (1992a). *Springs of Action: Understanding Intentional Behavior*. New York: Oxford University Press.

---------- (1992b). "Intentions, Reasons, and Beliefs: Morals of the Toxin Puzzle." *Philosophical Studies* 68:2, 171-194.

---------- (1995). "Effective Deliberation about What to Intend: Or Striking it Rich in a Toxin-Free Environment." *Philosophical Studies* 79, 85-93.

---------- (1996). "Internalist Moral Cognitivism and Listlessness." *Ethics* 106:4 (July 1996), 727-753.

Miller, Eric M., Gregory M. Walton, Carol S. Dweck, Veronika Job, Kali H. Trzesniewski, and Samuel M. McClure (2012). "Theories of Willpower Affect Sustained Learning." *PLoS ONE* 7(6): e38680.

Millikan, Ruth G. (1984). *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.

Milliken, John (2008). "In a Fitter Direction: Moving Beyond the Direction of Fit Picture of Belief and Desire." *Ethical Theory and Moral Practice* 11:5, 563–71.

Mintoff, Joe (2001). "Buridan's Ass and Reducible Intentions." *Journal of Philosophical Research* 26, 207-221.

Mischel, Walter (1996). "From Good Intentions to Willpower." In P. M. Gollwitzer and J. A. Bargh (eds.), *The Psychology of Action: Linking Cognition and Motivation to Behavior* (New York: Guilford Press), 197-218.

Montaigne (1877 [1580]). *Essays*. Tr. Charles Cotton. London: Reeves and Turner.

Montmarquet, James (1986), "The Voluntariness of Belief." *Analysis* 46:1, 49-53.

Moran, Richard (2001). *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.

Moore, G.E. (1942). "A Reply to My Critics." In Paul Arthur Schilpp, ed., The *Philosophy of G.E.Moore* (Chicago: Northwestern), 535-677.

---------- (1993). "Moore's Paradox." In Thomas Baldwin, ed., *G.E. Moore: Selected Writings* (New York: Routledge), 207-212.

---------- (1944). "Russell's Theory of Descriptions." In Paul Arthur Schilpp, ed., The Philosophy of Bertrand Russell (New York: Tudor), 175-225.

Muraven, M. (1998). *Mechanisms of Self-Control Failure: Motivation and Limited Resources*. Ph.D. dissertation, Case Western Reserve University.

Muraven, Mark and Roy F. Baumeister (2000). "Self-Regulation and Depletion of Limited Resources: Does Self-Control Resemble a Muscle?" *Psychological Bulletin* 126:2, 247-259.

Muraven, Mark, Dianne M. Tice, and Roy Baumeister (1998). "Self-Control as Limited Resource: Regulatory Depletion Patterns." *Journal of Personality and Social Psychology* 74:3, 774-789.

Navon, David (1984). "Resources—A Theoretical Soup Stone?" *Psychological Review* 91:2, 216-234.

Neurath, Otto (1983). "The Lost Wanderers of Descartes and the Auxiliary Motive." In Otto Neurath, *Philosophical Papers 1913-1946* (Dordrecht: D. Reidel Publishing Company), 1-12.

Nordgren, Loran F., Frenk van Harreveld, and Joop van der Pligt (2009). "The Restraint Bias: How the Illusion of Self-Restraint Promotes Impulsive Behavior." *Psychological Science* 20:12, 1523-9.

Olson, Karen (2007). "A New Way of Thinking about Fatigue: A Reconceptualization." *Oncology Nursing Forum* 34:1, 93-99.

Owens, David (2002). "Epistemic Akrasia." *The Monist* 85:3, 381-397.

Pears, David (1982). "How Easy is *Akrasia*?" *Philosophia* 11, 33-50.

Plato (1997). *Plato: Complete Works*. Ed. John M. Cooper. Indianapolis: Hackett.

Platts, Mark (1979). *Ways of Meaning*. Boston: Routledge and Kegan Paul.

Price, H.H. (1965). "Belief 'In' and Belief 'That'." *Religious Studies* 1:1, 5-27.

---------- (1969). *Belief*. London: Allen & Unwin.

Price, Huw (1989). "Defending Desire-As-Belief." *Mind* 98:389, 119–27.

Priest, Graham (2006). *In Contradiction* (2nd edition). Oxford: Clarendon Press.

Putnam, Hilary (1975). *Mind, Language, and Reality*. New York: Cambridge University Press.

---------- (1988). *Representation and Reality*. Cambridge, MA: The MIT Press.

Quine, W.V.O. (1956). "Quantifiers and Propositional Attitudes." *The Journal of Philosophy* 53, 177–186.

Quinn, Warren (1995). "Putting Rationality in its Place." In Rosalind Hursthouse, Gavin Lawrence, and Warren Quinn, eds., *Virtues and Reasons: Philippa Foot and Moral Theory* (New York: Oxford University Press), 181-208.

Raz, Joseph (1986). *The Morality of Freedom*. Oxford: Clarendon Press.

---------- (1997). "Incommensurability and Agency." In Chang (1997), 110-128.

---------- (2010). "On the Guise of the Good." In Tenenbaum (2010), 111-137.

Rescher, Nicholas (1959). "Choice without Preference." *Kant-Studien* 51, 142-175.

Roberts, John (2001). "Mental Illness, Motivation and Moral Commitment." *The Philosophical Quarterly* 51:202, 41-59.

Rorty, Amelie (1983). "Akratic Believers." *American Philosophical Quarterly* 20:2, 175-183.

Sartre, Jean-Paul (1957 [1946]). "Existentialism." ["L'Existentialisme est un humanisme."] In Jean-Paul, Sartre, *Existentialism and Human Emotions* (New York: The Wisdom Library), 9-51.

Scanlon, Thomas (1998). *What We Owe to Each Other*. Cambridge, MA: Belknap Press of Harvard University Press.

Schmeichel, Brandon J., Kathleen D. Vohs, and Roy F. Baumeister (2003). "Intellectual Performance and Ego Depletion: Role of the Self in Logical Reasoning and Other Information Processing." *Journal of Personality and Social Psychology* 85:1, 33-46.

Schwitzgebel, Eric (2002). "A Phenomenal, Dispositional Account of Belief." *Noûs* 36:2,

249-75.
---------- (2001). "In-Between Believing." *The Philosophical Quarterly* 51:202 (Jan. 2001), 76-82.
---------- (2010). "Acting Contrary to Our Professed Beliefs or The Gulf Between Occurrent Judgment and Dispositional Belief." *Pacific Philosophical Quarterly* 91, 531-53.
Schueler, G. F. (1991). "Pro Attitudes and Direction of Fit." *Mind* 100:2, 277–81.
Searle, John (1983). *Intentionality*. Cambridge: Cambridge University Press.
Setiya, Kieran (2007). *Reasons without Rationalism*. Princeton: Princeton University Press.
---------- (2008). "Believing at Will." *Midwest Studies in Philosophy* 32, 36-52.
---------- (2010). "Sympathy for the Devil." In Tenenbaum (ed.), *Desire, Practical Reason, and the Good* (New York: Oxford University Press), 82-110.
Shah, Nishi (2003). "How Truth Governs Belief." *The Philosophical Review* 112, 339-93.
---------- (2008). "How Action Governs Intention." *Philosophers' Imprint* 8:5, 1-19.
Shah, Nishi and David Velleman (2005). "Doxastic Deliberation." *The Philosophical Review*, 114:4 (Oct. 2005), 497-534.
Sherkoske, Greg (2010). "Direction of Fit Accounts of Belief and Desire Revisited." *Croatian Journal of Philosophy* 28: 1-11.
Shoemaker, Sydney (1995). "Moore's Paradox and Self-Knowledge." *Philosophical Studies* 77:2/3, 211-228.
Simplicius (1984). *Simplicii in Aristotelis de Caelo Commentaria*. In Ed. I. L. Heiberg, ed., *Commentaria in Aristotelem Graeca*, vol.7. (Berlin: Royal Prussian Academy).
Smith, Michael (1994). *The Moral Problem*. Oxford: Blackwell.
Sobel, David, and David Copp (2001). "Against Direction of Fit Accounts of Belief and Desire." *Analysis* 61:1, 44–53.
Sorensen, Roy (1988). *Blindspots*. New York: Oxford University Press.
---------- (2007). "The All-Seeing Eye: A Blind Spot in the History of Ideas." In Green and Williams (2007), 37-49.
Spinoza (2002). *Complete Works*. Tr. Samuel Shirley. Indianapolis: Hackett.
Stocker, Michael (1979). "Desiring the Bad." *The Journal of Philosophy* 76: 738-53.
Strickland, Lloyd (2006). "God's Problem of Multiple Choice." *Religious Studies* 42:2, 141-57.
Stöltzner, Michael M. (2000). "An Auxiliary Motive for Buridan's Ass: Otto Neurath on Choice without Preference in Science and Society." *Conceptus* 33:82, 23-44.
Svavarsdóttir, Sigrún (2006). "How Do Moral Judgments Motivate?" In James Dreier, ed., *Contemporary Debates in Moral Theory* (Oxford: Blackwell), 163-181.
Taylor, Kenneth A. (2002) "De re and de dicto: Against the Conventional Wisdom." *Philosophical Perspectives* 16, 225–265.
Tenenbaum, Sergio (2007). *Appearances of the Good: An Essay on the Nature of Practical Reason*. New York: Cambridge University Press.
---------- (2009). "Knowing the Good and Knowing What One is Doing." *Canadian Journal of Philosophy* 39 (supplement), 97-119.

----------, ed (2010). *Desire, Practical Reason, and the Good*. New York: Oxford University Press.

---------- (2014). "Minimalism about Intention: A Modest Defense." *Inquiry* 57:3, 384-411.

Ullmann-Margalit, Edna, and Sidney Morgenbesser (1977). "Picking and Choosing." *Social Research* 44:4, 757-785.

Van Fraassen, Bas (1984). "Belief and the Will." *The Journal of Philosophy* 81, 235-56.

Velleman, J. David (1989). *Practical Reflection*. Princeton, NJ: Princeton University Press.

---------- (2000a). "The Guise of the Good." Reprinted in J. David Velleman, *The Possibility of Practical Reason* (Oxford: Oxford University Press), 99-122.

---------- (2000b). "Introduction." In J. David Velleman, *The Possibility of Practical Reason* (Oxford: Oxford University Press), 1-31.

Vohs, Kathleen D. and Roy F. Baumeister (2004). "Ego Depletion, Self-Control, and Choice." In Jeff Greenberg, Tom Pyszczinsky, and Sander L. Koole, eds., *Handbook of Experimental Existential Psychology* (New York: Guilford Press, 2004), 398-410.

Watson, Gary (1977). "Skepticism about Weakness of Will." *The Philosophical Review* 86:3, 316-339.

Williams, Bernard (1970). "Deciding to Believe." In Bernard Williams, *Problems of the Self* (New York: Cambridge University Press), 136-51.

Williams, John N. (1992). "Belief-in and Belief in God." *Religious Studies* 28:3, 401-406.

Williamson, Timothy (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.

Winters, Barbara (1979). "Believing at Will." *The Journal of Philosophy* 76:5, 243-56.

Wittgenstein, Ludwig (2001 [1953]). *Philosophical Investigations*, 2nd edition. Tr. Elizabeth Anscombe. Oxford: Blackwell.

---------- (1980a). *Remarks of the Philosophy of Psychology*, volume I. Ed. G.E.M. Anscombe and G.H. von Wright. Chicago: University of Chicago Press.

---------- (1980b). *Remarks of the Philosophy of Psychology*, volume II. Ed. G.H. von Wright and Heikki Nyman. Chicago: University of Chicago Press.

Zangwill, Nick (1998). "Direction of Fit and Normative Functionalism." *Philosophical Studies* 91:2, 173–203.