A Deontological Explanation of Accessibilism

by
Michael Hatcher

A Dissertation Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements of the Degree
DOCTOR OF PHILOSOPHY
(PHILOSOPHY)

May 12th 2017

ProQuest Number: 10268338

ProQuest 10268338

TABLE OF CONTENTS

Acknowledgments

There are many people to thank.

Deep thanks are owed to my dissertation advisor, Ralph Wedgwood. As an undergrad in 2007, I attended Ralph's lecture series in epistemology during a semester I had overseas in Oxford. It was my first serious introduction to epistemology, and his bold yet persuasive approach to the discipline's most central questions inspired me. I had no idea that, years later, I would be lucky enough to have him as my advisor at USC. Ralph holds my writing to a high standard. I still remember how he poked a plethora of holes through the first paper of mine we discussed. At the same time, his patience and kindness continually gave me confidence to try again, to see what could be learned from his criticisms, and to develop my ideas more clearly and fully. Though this dissertation surely contains faults, Ralph's guidance has made it far better than it otherwise would have been.

I am grateful for the members of my dissertation committee. Jim Van Cleve regularly clued me in to crucial distinctions and nuances. And without a hint of rushing, I always felt he enjoyed exploring philosophical space with me. I loved my discussions with Janet Levin about attention and phenomenal consciousness, and how she encouraged me to stick to my guns. More times than I can count, Mark Schroeder's strategic insight helped me better see where my efforts in philosophy were likely to bear fruit. And I thank Greg Keating for challenging my ideas about the conditions for blameworthiness.

Many thanks are owed to my teaching mentor, Shieva Kleinschmidt. Both in my capacity as one of her TA's and by meeting with her, I learned from her that much can be done to prepare for teaching a class. Full of insight and inspiration, she is a fantastic mentor.

I would also like to thank Susan Harris, my advisor at the Joint Educational Project at USC. Besides her cheerful and wise guidance concerning all things JEP, Susan went above and beyond the call of duty to help me secure paternity leave for Spring 2017.

With perception and assiduity, Steve Finlay and Gideon Yaffe helped me improve my first few papers at USC. And throughout, I have had kind, engaging colleagues. Matt Babb, Kenneth Silver, Nick Laskowski, and Abelard Podgorski have been particularly helpful comrades in doing philosophy. There are others to thank, perhaps too many to name: Mike Ashfield, Maegan Fairchild, Ara Astourian, Matt Leonard, Alex Dietz, Michael Pressman, Caleb Perl, Jonathan Wright, Aness Webster, Steve Bero, Michael Milona, Aaron Veek, Tanya Kostochka, Woo Ram Lee, Matt Lutz, Rima Basu, Renee Bolinger, Shyam Nair, Jesse Wilson, Kenny Pearce, Lewis Powell, Joe Horton, Alida Liberman, and Justin Dallmann.

Before USC, I learned under David Pitt and Mark Balaguer at California State University, Los Angeles. I cannot thank David and Mark enough for the time they invested in me. They helped me build theories while patiently and kindly forcing me to write more clearly. And while I am thinking of CSULA, I want to shout-out to friends and colleagues who made my time there thoroughly enjoyable: Matt Leonard, Zeph Scotti, Michael Anderson, and Elton Hollon.

My first exposure to philosophy was as an undergrad at Biola University. At Biola, I learned under teachers who had the posture that God enjoys doing philosophy with us, and is pleased to share with us whatever insights are helpful to us at the time. Gregg Ten Elshof, Tom Crisp, and David Ciocchi saw themselves as servants and intercessors for their students, and their overflowing kindness left on me a lasting impression of how philosophy could be pursued as part of a whole, flourishing life.

I have spent seven years at USC. These years were marked with difficult transitions. My parents divorced, one of my brothers went through a divorce, and my dad was diagnosed with cancer. My community, though, rallied around me to help me process my feelings, avoid unhelpful patterns of thought, and regain joy in moving forward. My pastor and mentor Beau Watkins has always been a phone call away. John Kwak and Justin Goll have been better friends than I could ask for. And I was lucky to have had a handful of conversations with Dallas Willard before he passed away in 2013. Each conversation lifted a load off my back. He had an uncanny perceptiveness of the terrain my soul was in, and helped me see where to step next. At around this time, InterVarsity pastor Joe Thackwell invited me into a spiritual formation reading group of fellow USC doctoral students. Discussing the spiritual life with Jared Warner, Peter Meindl, and Mark Harrison proved not only encouraging, it began friendships I treasure. Later, Joe connected my wife and I to Jane Willard in her capacity as a spiritual director. As our director, Jane has been a stabilizing anchor for us. She helps me see the many places Jesus' love touches my life.

My deepest thanks are for my family: my parents, my brothers, my parents-in-law, and my siblings-in-law. I cannot measure or adequately describe how much each of you mean to me. I do not know where to begin. But in any case, perhaps it is best that I tell you in person. Most of all, I thank you Katie, my love, for your strong mind and your big, kind, and good heart. You are a continual source of joy for me. And though you are too young to read this right now, thank you Baby Clare for expanding that joy.

Introduction

Epistemic internalism is the view that epistemic rationality is determined by what is internal to the subject. On a historically prominent version of internalism, rationality is determined by what is accessible to the subject. What is accessible to the subject is usually thought of as what is available to her consciousness. A common explanation for why accessibilism should be true runs as follows: it is rational for one to believe a proposition just in case it is blameless for one to do so; but whether it is blameless for one to do so is determined by what is accessible to one; so, rationality is determined by what is accessible to one.

Let us call *accessibilism* the version of internalism on which rationality is determined by what is accessible to the subject. And let us call *deontologism* the view that it is rational for one to believe a proposition just in case it is blameless for one to do so. The explanation of accessibilism just mentioned is a *deontological* explanation: it appeals to deontologism. Accessibilism has fallen on hard times. And so have deontological explanations of accessibilism.

Several objections to deontological explanations are responsible for these hard times. Chief among them is the following. Deontologism implies we have epistemic obligations such that if we violate them, we are blameworthy. But this means belief is something one can control. For control is required if the question of whether blame is appropriate is to even arise. But we do not have control over belief. Instead, our beliefs are at the mercy of our evidence. So, deontologism is false. We shall call this the *At the Mercy of Evidence Objection*.

This objection poses a structural problem for deontological explanations of accessibilism. If we are to use deontologism to explain accessibilism, we need an understanding of epistemic obligations on which our having these obligations would explain why epistemic rationality is determined by what is accessible. And then we need to show that we have the kind of control over belief these obligations imply. Put differently, to address the At the Mercy of Evidence

Objection, we need to do two things. First, we need to develop a substantive explanation of why accessibilism should be true if deontologism is true. Second, we need to show that we really have the kind of control this explanation presupposes. This dissertation centers on accomplishing these two tasks.

To dive right into the core chapter of the dissertation, I appeal to action theory in Chapter 4 to establish the link between deontologism and accessibilism. Here is a sketch of the link. One is blameworthy only for exercises of control. But control always involves responsiveness to reasons. And *the person*, as opposed to one of her proper parts, is responsive to a reason only if she is conscious of it. Hence, one is blameworthy only for responses to reasons of which one is conscious. But consciousness is a paradigm form of accessibility. So, for a given way of responding to reasons, it is only with respect to *accessible* reasons that the question arises whether it is blameworthy or instead blameless to respond in that way. And now suppose deontologism is true: rationality is blamelessness. It follows that whether it is rational to respond in a given way to reasons is simply a matter of which reasons are accessible to one. And so accessibilism is true, too.

In Chapter 5, I argue that we should understand the control at issue in Chapter 4 as *the will*. A striking feature of the will is that it is exclusively responsive to reasons of which one is conscious. Moreover, there are independent reasons to think blameworthiness is tied to the will. In this way, Chapters 4 and 5 clarify what kind of control is presupposed by our having the relevant kind of epistemic obligation, and so clarify what it takes to reply to the At the Mercy of Evidence Objection. It takes showing that forming belief can be an exercise of will. Showing this is not easy. But, as I shall argue in Chapters 7-9, it can be done.

In Chapter 6, before addressing the At the Mercy of Evidence Objection head-on, I deal with an important preliminary worry to which Chapters 4 and 5 immediately give rise. Here is

the worry. It is unclear whether exercises of will are as ubiquitous as blameworthiness. In particular, it is unclear whether the forming of enduring states like intention and belief – for which we can be blameworthy, intuitively – can be exercises of will. In Chapter 6, I develop the idea that enduring states like these are grounded in *activities*. In developing this idea, I define activities, roughly, as events whose temporal parts can be exercises of will. With this idea on the table, we have a clearer idea of how to show that forming belief can be an exercise of will: show that belief is grounded in an activity. For then we can identify forming belief with the initial phase of the relevant activity.

In Chapter 7, I distinguish credence from outright belief, and then argue that credence, rather than being grounded in an activity, is actually a particular kind of feeling. Thus, forming a credence cannot be an exercise of will. Moreover, I argue that one's response to evidence is complete once one has a particular credence. If this is right, it must be that our epistemic obligations aren't simply to respond to evidence in particular ways, or, put differently, it must be that epistemic reasons include not only evidence but also practical reasons. In Chapter 8, I argue that outright belief, by contrast with credence, is grounded in an activity. On my account, outright belief is grounded in a particular way of organizing one's attention. With this account on the table, I argue, in Chapter 9, that evidence does not exhaust the epistemic reasons. As we shall see, part of what shows this is that, when it comes to outright belief, there is more to correctness than truth.

Chapters 7-9 do heavy-lifting to defang with the At the Mercy of Evidence Objection. Fortunately, as I show in Chapter 10, this heavy-lifting allows us to dispatch with another objection often pressed against deontological explanations of accessibilism. We can call this objection *the Dilemma*: Either accessibility involves belief, or it does not. If it does, then it makes sense why reasons should be accessible. For our beliefs affect what it is blameworthy for

us to do. But then forming a belief is rational only if other beliefs are in play, and so we have to give up on foundationalism. On the other hand, if accessibility does not involve belief, it is not clear why reasons should be accessible in the first place. For it is not clear how what we do not so much as believe affects what it is blameworthy for us to do. According to the Dilemma, deontological explanations either lose their integrity, or else conflict with foundationalism. As I said, Chapters 7-9's heavy lifting allows us to dispatch with this objection. For Chapters 7-9 result in an overall theory on which epistemic obligations apply to outright beliefs, but not credences. But then reasons accessible to us via credences affect what it is blameworthy for us to do. So, it makes sense why reasons should be accessible. At the same time, we can preserve foundationalism, so long as it is restricted to outright belief. In this way, the different theoretical roles of credence and outright belief dissolves the Dilemma.

In this way, in Chapters 4-10, I develop a substantive deontological explanation of accessibilism, respond to the At the Mercy of Evidence Objection, and respond to the Dilemma. Of course, the At the Mercy of Evidence Objection and the Dilemma are not the only reasons why deontological explanations of accessibilism have fallen on hard times.

Two other problems are worth mentioning. The first problem: some have argued that even if deontologism provides some traction for endorsing accessibilism, accessibilism itself generates an infinite regress of facts that must be accessible to one, if one is ever to be rational. But that is absurd. In Chapter 2, I use this regress objection as an opportunity to clarify accessibilism. In fact, I argue that accessibilism is ambiguous, that only one disambiguation is regress-inducing, and that the regress-free disambiguation is better motivated by the intuitive pressures by virtue of which accessibilism exists in the first place. The second problem: others have argued, quite persuasively, that past attempts to generate traction from deontologism have failed. This generates pessimism about deontological explanations. In Chapter 3, I show that

attention to why past attempts have failed allows us to better understand what deontologism would have to look like, if it is to explain accessibilism. In this way, in Chapters 2 and 3, I treat these two problems as opportunities to clarify accessibilism and deontologism. These clarifications are important, as they make more precise the very structure a deontological explanation would need to have, to be successful. So, they are a natural place to start, before giving my substantive explanation in Chapter 4.

Indeed, only the very beginning is a more natural place to start. Chapter 1 starts at the very beginning. It offers an introduction to accessibilism, deontologism, the intuitive pressures from which they arise, and the problems deontological explanations face.

I end this introduction with a few remarks. I want to make clear, from the start, that my dissertation includes no substantive argument for deontologism. There are reasons for this absence besides considerations of space. Deontologism is a view about epistemic rationality. At a minimum, we can think of epistemic rationality as a kind of rationality unique to, or suited to, belief. But even so, there are many different senses of the word 'rational'. There is a sense in which one with probabilistically incoherent credences is irrational, even if one never could have prevented the incoherence. But there is also a sense in which such an agent might well be perfectly rational, and just unfortunate. Moreover, without a concrete crystallization of deontologism's presuppositions, it is hard to pin down whether a particular use of 'rational' accords with the deontologist's sense. For both of these reasons, it is unclear what kind of task one is taking on, if one attempts a straightforward argument that epistemic rationality is to be understood in terms of blamelessness. But all is not lost. Once my deontological picture is on the table, it will be possible to look and see that 'rational' and 'irrational' appear, at least in some contexts, to be used in the manner the picture predicts. This procedure gives us a firm grip on a *kind* of epistemic rationality.

In this connection, I do not presuppose, as some do, that rationality is worth calling 'epistemic' only if it is an essential ingredient of knowledge. For one, even focusing on 'know that' phrases, there are different senses of 'know'. Plausibly, each of the following senses differ: the sense in which the automatic door knows that a person is approaching, the sense in which the birds know that they need to fly to Maine for the summer months, and the sense in which you know that you have a safe, reliable car. For this reason, I suspect that tying epistemic rationality to knowledge is premature, as it either unduly restricts the role of epistemic rationality, or else promotes confusion. By my lights, it is better to firm up a notion of a kind of epistemic rationality first, and then ask how that kind may relate to various kinds of knowledge. I do not tackle the latter task. But this dissertation can be seen as a propaedeutic for it.

OUTLINE

Chapter 1: Problems for a Deontological Explanation of Accessibilism

§1.1   Introduction

Consider the following line of thought. In one sense of the word 'rational', to say it is

epistemically rational for you to believe something is just to say that no one should hold it

against you if you did. But whether it is appropriate to blame you is a matter of whether you

responded to the situation you found yourself in as well as could be reasonably expected of you.

Now, no one should hold it against you if you failed to take into account in your reasoning, and

so respond to, a reason to which you had no access in the first place. Thus, what is accessible to

you settles the matter of whether it is appropriate to blame you. But remember that rationality is

just blamelessness. Therefore, to conclude the line of thought, whether it is rational for you to

believe something is determined by what is accessible to you.

The above paragraph is philosophically un-careful. Distinctions and clarifications will

need to be made if it is to survive scrutiny. But it does qualify as a statement of what I shall call

a deontological explanation of accessibilism. The view that it is epistemically rational to believe

just in case it is blameless to do so is often called *epistemic deontologism* – or, for short,

deontologism. That is where the line of thought began. And the view that rationality is

determined by what is accessible to the subject is *accessibilism*. That is where the line of thought

ended, and it is the view the line of thought was an attempt to *explain*.

My project in this dissertation is to develop and defend a deontological explanation of

accessibilism. According to Alvin Plantinga,[1] such explanations trace back to Descartes and

Locke. As Plantinga sees it, Descartes and Locke each endorse accessibilism on the basis of two

views: that epistemic justification or rationality consists in its being blameless for one to believe,

---

[1]   Plantinga 1993, 11-25.

and that whether it is blameless to believe is determined by what is accessible. The first view is deontologism, which, combined with the second, implies accessibilism. Descartes and Locke are also the figureheads of *internalism*, i.e., the view that rationality is determined by what is, in some sense, *internal* to the subject. Indeed, I shall take it as a constraint on our understanding of accessibilism that it is a version of internalism.

Since a paper by Earl Conee and Richard Feldman, though, philosophers have been keen to distinguish accessibilism from *mentalism*, which is another version of internalism.[2] Mentalism is the view that rationality is determined by the subject's mental states. Descartes and Locke endorse both accessibilism and mentalism. But, importantly, they endorse mentalism *because* they endorse accessibilism – the connecting thought being, for them at least, that only mental states are accessible in the relevant way. In any case, if Plantinga is right about Descartes and Locke, it is perhaps fair to conclude that the driving force behind internalism in the early modern period is a deontological explanation of accessibilism.

As Richard Schantz has said, the debate over internalism is a debate about "the very form that theories in epistemology…ought to take."[3] However, developing a deontological explanation of accessibilism should also impact extra-epistemological concerns. For example, several have noticed that, in many cases, *moral* blameworthiness appears to depend entirely on *epistemic* blameworthiness.[4] A developed deontological explanation would help us see what it might take for deontologism to be true, and so whether or not one could be blameworthy, in the first place, for forming a belief. This may help us see whether it could be that moral blameworthiness sometimes depends entirely on epistemic blameworthiness.

---

[2]   Conee and Feldman 2001.
[3]   Schantz 2004, 1.
[4]   See Nottelmann (2007, 3-10) and Montmarquet (1993, 1-10).

The dissertation's driving impetus, though, is not possible downstream applications of a deontological explanation, but rather the question of whether this kind of explanation *could* work, in the first place. Many objections have been raised against this type of explanation, and it is often treated as dead in the water.

In this chapter, I have two main goals. My first goal, in §1.2, is to clarify in more detail what a deontological explanation of accessibilism is supposed to be, in the first place. This requires some preliminary discussion of epistemic rationality, accessibilism, and deontologism. My second goal, in §1.3, is to clarify some of the problems that face deontological explanations of accessibilism. The problem which plays the most significant role in this dissertation is one I shall call the *At the Mercy of Evidence Objection*. According to this objection, belief is at the mercy of evidence; hence, we do not have the control over belief we would need to have for deontologism to be true. I shall also discuss three other problems for deontological explanations. I shall conclude, in §1.4, by explaining the method behind the sequence in which I shall address each of these four problems in coming chapters.

§1.2   What is a Deontological Explanation of Accessibilism?

What is a deontological explanation of accessibilism? Let us break this question into four simpler questions. Deontologism and accessibilism are views about epistemic rationality. So the first question is: what is *epistemic rationality*? Second, what is *accessibilism*? Third, what is *deontologism*? And fourth and finally, what kind of *explanation* do I have in mind, when I say that my project is to develop a deontological explanation of accessibilism?

§1.2.1 Epistemic Rationality

Epistemic rationality is a kind of rationality that has to do with *doxastic* states or attitudes, i.e., belief, suspension, denial, and credences. For brevity, I shall often talk simply of

belief. I shall assume that a kind of rationality is epistemic only if, at a minimum, it is appropriate or unique to belief. What makes a kind of rationality appropriate or unique to belief? Suppose that you are offered a large bribe to believe it is raining in Riverside. Perhaps it is rational to believe it is raining in Riverside in *some* sense of 'rational'. But this is the same sense in which it is rational to eat a gallon of ice cream if offered a bribe to do so. Such rationality is not appropriate or unique to belief. By contrast, if one has solid evidence that it is raining in Riverside, it might well be rational to believe it is raining in a sense of 'rational' appropriate or unique to belief. Now, the question of what it is for rationality to be appropriate or unique to belief turns out to be a large one, to which I return in §9.2.

For now, I shall clarify some things I will *not* assume about epistemic rationality. I shall not assume there is only one kind of rationality appropriate or unique to belief. For example, the sense in which an agent with probabilistically incoherent credences is irrational seems appropriate or unique to belief, even if she could never have prevented the incoherence. But there may also be a kind of rationality appropriate or unique to belief such that this agent might well be perfectly rational, and just unfortunate. So, there may be more than one kind of epistemic rationality, suited to different theoretical purposes.

In addition, I shall not assume that epistemic rationality is an essential ingredient of knowledge. This contrasts with other ways people have demarcated epistemic rationality. For example, John Greco understands epistemic justification to be "the normative dimension of knowledge".[5] And Mark Schroeder defines epistemic rationality as the strongest kind of rationality entailed by knowledge.[6] One reason I withhold from this kind of assumption about epistemic rationality is that, even focusing on 'know that' phrases, there are different senses of

---

[5]   John Greco 2010, 47, especially footnote 1.
[6]   Schroeder 2012b, 268.

'know'. Plausibly, each of the following senses differ: the sense in which the automatic door

knows that a person is approaching, the sense in which the birds know that they need to fly to

Maine for the summer months, and the sense in which you know that you have a safe, reliable

car. Hence, tying epistemic rationality to knowledge promotes confusion, unless an independent

clarification of knowledge is in hand. But one might think that the possible kinds of epistemic

rationality are among the tools we need, to offer such a clarification. Or else, if tying epistemic

rationality to knowledge does not promote confusion, it may unduly restrict the role of epistemic

rationality. For if having knowledge is what is in common between the automatic door, the bird,

and you, then whatever kind of epistemic rationality is essential to knowledge begins to look

rather attenuated, indeed.

My approach is to assume that all kinds of epistemic rationality – and all kinds of

knowledge, for that matter – are independently worthy of theoretical investigation and

development. By my lights, it is just sound methodology to begin postulating relationships

between a kind of rationality and a kind of knowledge only *after* the relevant kinds have been

theoretically developed and clarified.

§1.2.2 Accessibilism

According to Jessica Brown,

> Epistemic internalism is the view that a thinker's epistemic status depends wholly
> on matters which are 'internal' to that thinker, rather than at least partially on
> matters which are 'external' to her, such as her relations to her environment.[7]

Of course, knowledge is an epistemic status. But since knowledge implies truth and truth does

not depend wholly on what is internal to the subject, it is standard to restrict internalism to

justification or, on the terminology I prefer, rationality. It is also common to understand what is

---

[7]   Brown 2007, 13-14.

internal to the subject as what is accessible to her. For example, on the view Robert Audi calls

'internalism about justification',

> …justification is grounded entirely in what is internal to the mind, in a sense
> implying that it is accessible to introspection or reflection by the subject…[8]

This common form of internalism is known as *accessibilism*.

We shall gloss the view as follows:

accessibilism   Whether believing *p* is rational for S is determined by what is accessible to
S.

I shall discuss in more detail how to best define accessibilism in Chapter 2. For now, I should

mention that the above gloss is silent on several distinctions that will be important later in the

dissertation. For example, it is silent on the distinction between its being rational to believe *p* and

rationally believing *p*, and the distinction between believing as having a belief state and

believing as forming a belief state.

To better understand accessibilism, it is important to see why it exists in the first place –

that is, to understand the intuitive pressures or motivations from which it arises. We can begin

with Laurence BonJour's reply to D.M. Armstrong. An early proponent of externalism in

epistemology, Armstrong held that "what makes a true non-inferential belief a case of

knowledge" – and so, presumably, also *rational* – is a lawlike connection between the belief and

what makes the belief true.[9] Armstrong compared this connection to that between a

thermometer's mercury level and the temperature. In his reply to Armstrong, BonJour gives the

example of Norman the clairvoyant.[10] A lawlike connection holds between Norman's belief that

the President is in NYC and the President's being in NYC, though no reasons for or against this

---

[8]   Audi 1998, 233-234.
[9]   Armstrong 1973, 157.
[10]   BonJour 1985, 41-44.

belief are among what is accessible to him. Intuitively, his belief is not rational, contra

Armstrong.

BonJour asks:

> …*why* should the mere fact that such an external relation [i.e., the lawlike
> connection] obtains mean that Norman's belief is epistemically justified when the
> relation in question is entirely outside his ken?[11]

This rhetorical question drives toward the general idea that only what is accessible to the subject

helps determine rationality.

Next to the case of Norman, consider what is known as the *new evil demon problem* for

reliabilism.[12] On reliabilism, a belief is rational just in case it is formed by a belief-producing

mechanism reliable at producing true beliefs. Assume many of our belief-producing mechanisms

are reliable in the actual world. Now consider a possible world in which we undergo the exact

same experiences, processes of reasoning, and form exactly the same beliefs we do in the actual

world except that, unbeknownst to us, a demon ensures our experiences systematically mislead

us. Reliabilism implies that most of our beliefs in the actual world are rational, but that most of

our beliefs in the demon world are not rational. However, intuitively, the same beliefs are

rational in both worlds. Interestingly, what is accessible to us appears to be the same in each

world. Moreover, just as BonJour said of Norman that his lawlike connection with the truth does

not make him rational because what is accessible to him does not include it, there is temptation

to say that the demon's interference would not make us irrational because what is accessible to

us would not include it. We can hear the echo of the idea that only what is accessible to the

subject helps determine rationality.

---

[11]   Ibid., 42.
[12]   See Ball and Blome-Tillmann (2013) and Wedgwood (2002a, 349). Cohen (1984, 281-284) is an early source. The *old* evil demon
problem is the skeptical problem Descartes put forth in his *Meditations*.

Accessibilists often link the intuitions we tend to have about Norman and the new evil demon problem to deontologism, on which rationality consists in blamelessness. If we combine deontologism with the intuitive idea that whether it is blameless for one to believe is determined by what is accessible to one, we can explain why these intuitions should be taken seriously. I shall say more about deontologism in the next section. And, of course, this entire dissertation is about whether this kind of appeal to deontologism can work.

It is worthwhile making a few points about what could qualify as an accessibilist view. Consider, for example, Timothy Williamson's view that the rationality of one's beliefs is determined by what one knows.[13] If we thought of what one knows as what is accessible to one, then this would qualify as an instance of accessibilism. But here is where the juxtaposition of the normal and the demon world can serve as a kind of touch-point. I shall assume accessibilism is a version of *internalism* in epistemology. Hence, we must understand what is accessible to one as *constant* between the normal world and the demon world. For it is only external things that vary between these worlds. Since what one knows varies between these worlds, what one knows is not one and the same thing as what is accessible to one. And so Williamson's view does not qualify as a version of accessibilism.

In this connection, accessibility is commonly understood in terms of knowability by reflection alone. If we think of reflection as any combination of introspection and *a priori* reasoning, then what is knowable by reflection alone is constant between the normal and the demon worlds. I shall assume that reflection, so understood, is a kind of *consciousness*. And indeed, in general, I shall assume that for something to be accessible, a certain kind of

---

[13]   Williamson 2000, 191.

consciousness of it must be available, if not actual. However, I shall not assume that accessibility must be understood as knowability by reflection alone.

§1.2.3 Deontologism

The word 'deontologism' comes from a Greek word for 'duty'. It is worth keeping clear from the start, though, that in the context of 'epistemic deontologism', 'deontologism' refers neither to the view typically contrasted with consequentialism in ethics, nor to an epistemic analogue of that view. Deontologism is, indeed, sometimes formulated in terms of 'duty', but despite this nominal, etymological overlap with the view contrasted with consequentialism, deontologism involves no commitments about the relationship between rationality and consequences or outcomes.

As it is variously formulated, deontologism is the view that epistemic rationality consists in *fulfilling one's duties*, *fulfilling one's obligations*, *believing as one ought*, or *blamelessness*.[14] It is worth pausing briefly over these different notions. Fulfilling duties and obligations can be understood as believing as one ought, at least in some sense of 'ought'.[15] But there are many senses of 'ought'. For example, in one sense, we ought to believe all the infinitely many true propositions, and not believe any of the infinitely many false propositions. But one is not necessarily blameworthy for failing to believe as she ought in this austere and idealistic sense of 'ought'. Moreover, believing as one ought, in this idealistic sense, seems disconnected, or at least a step removed, from our intuitive notion of rationality. So what sense of 'ought' is at issue?

---

[14] See Pryor (2001, 111) and Bergmann (2006, 77) for formulations of epistemic deontologism.

[15] Throughout this dissertation, 'S ought to ___' and 'S has an obligation to ___' are treated as synonymous. We might observe that 'S ought to A' might be true even though 'S has an obligation to A' is false, and the idea is that 'ought' points out which courses of action are *maximizing*, while 'obligation' points out which courses of action are *required*. I have no problem with this idea, though I also see no reason to deny that a *sense* of 'ought' might be requiring instead of maximizing. Indeed, the literature I discuss tends to treat 'ought' as a synonym of 'obligation'. So, we must keep in mind that that is my practice, as well.

As an escape from this morass, I take the fact that 'blameless' almost always occurs in formulations of deontologism to indicate that the notion of blamelessness, or avoiding appropriate blame, is a notion in terms of which we can understand each of the others.

Of course, in one sense of 'blameless', something is blameless even when no obligations apply to it in the first place, as when we say that we cannot blame the stars for twinkling as they do, or my stomach for digesting food as it does. In this dissertation, we shall never understand blamelessness in this trivial or degenerate way. Instead, blamelessness is always the fulfilling of an obligation. Moreover, in some sense of 'can', an obligation applies to one only if there is both something blameworthy one *can* do at the time as well as something blameless one *can* do at the time instead. Then, to mark conceptual connections between blamelessness, blameworthiness, and obligation: blamelessness is avoiding the blameworthy thing, blameworthiness is avoiding the blameless thing, and one's obligation is to do the former rather than the latter. And so we can understand the sense of 'ought' at issue, when deontologism is construed in terms of believing as one ought. It is the sense such that if one does what one ought not, one is blameworthy, just as the relevant kind of duty or obligation is that whose violation implies one is blameworthy.[16]

With this clear, we can give the following initial gloss on deontologism:

deontologism   Believing *p* is rational for S if and only if believing *p* is blameless for S. Like our gloss on accessibilism, this formulation of deontologism is silent on several distinctions that will be important later on.

As was the case with accessibilism, to better understand deontologism, it is important to understand the intuitive pressures or motivations from which it arises. The first thing to see is that there appears to be at least a *sense* of the word 'rational' such that believing's being rational

---

[16]   In understanding deontologism in this way, I follow Nikolaj Nottelmann 2007.

for one consists in its being blameless for one. Consider someone's believing that the earth is flat or, for an example perhaps closer to home, that not vaccinating one's children better promotes health for all children. It is easy to imagine one uttering 'That's irrational!' to such a person, in the characteristic blaming tone of voice. In this context at least, irrationality is blameworthiness, and, hence, rationality is a certain kind of blamelessness.

This first point about the term 'rational' already brings out that we do, in fact, blame people for believing as they do. Moreover, there is a reason to think it is, in fact, appropriate to blame people in this way. For it appears that certain intuitively blameworthy actions or omissions can be explained only if the subject is blameworthy for believing as she does. Consider, for example, W.K. Clifford's famous example of the Shipowner.[17] The Shipowner has some doubts whether his aging vessel is still sea-worthy. But he ignores these doubts, and reminds himself how much inductive evidence he has of the old ship's sea-worthiness. He comes to believe the ship is safe, and so omits refitting the hull. Subsequently, the ship goes down and many lives are lost. The Shipowner is blameworthy for this omission, but he wouldn't be blameworthy if he were blameless for believing that the ship is safe. So, he is, in fact, blameworthy for believing this.[18]

So, it looks like, in at least one sense of the word, 'irrational' expresses blame directed at beliefs. And this blame appears very often appropriate. These facts motivate deontologism, the view that epistemic rationality consists in blamelessness.

§1.2.4 What Kind of Explanation Am I After?

In many contexts, we are perfectly content with explanations[19] that are partial in the sense

---

[17]   Clifford 1877, 289-290.
[18]   For similar examples, see Nottelmann (2007, 5-6) and Montmarquet (1993, 1-2).
[19]   What is an explanation of a fact, F? The word 'explanation' either refers to a set of facts by virtue of which F obtains, or instead to a set of statements that cite such a set of facts. This ambiguity is perfectly intelligible and mostly harmless, though, because it is easy to see through on the basis of context.

that they cite facts that do not entail what they explain. Consider a case. I arrive home to a

flooded apartment. I find that I had left a faucet running. The fact that I had left a faucet running

does not entail the flooding of the apartment. My wife could have arrived home from work

unexpectedly, in time to stop the flood. A meteor could have leveled the entire apartment. And

so on. But that I had left the faucet running is an explanation of why the apartment flooded that I

find perfectly adequate.

However, the deontological explanation of accessibilism I am after shall take the

linguistic form of a valid argument. Since valid, the argument will express propositions which, if

true, entail accessibilism. So, the explanation I am after is not partial in the above sense.

Given our glosses on deontologism and accessibilism, i.e.,

deontologism Believing *p* is rational for S if and only if believing *p* is blameless for S.

accessibilism Whether believing *p* is rational for S is determined by what is accessible to
S.

the kind of explanation I am after must involve not only deontologism but also a bridge principle

we can gloss as follows:

the Bridge Principle Whether believing *p* is blameless for S is determined by what is
accessible to S.

Of course, a mere valid argument is a very low bar. Compare: BonJour has only true beliefs;

BonJour believes accessibilism; so, accessibilism is true. But this argument is no explanation at

all. Even a valid argument with premises that have a degree of plausibility is a fairly low bar.

Indeed, the conjunction of deontologism and the Bridge Principle already passes this bar. We

already discussed the intuitive support for deontologism. Now reflect on the Bridge Principle in

light of the new evil demon problem, or the case of Norman the clairvoyant. Since there is some

pull to the idea that blamelessness is determined by what is accessible to the subject, this second

bar is already passed.

I am after, though, an explanation of *why* the conjunction of deontologism and the Bridge Principle should be true. More precisely, I am after a story that, if true, cites facts *by virtue of which* this conjunction obtains. And I want to see if such a story *could*, in fact, be true. Now, for nearly any facts by virtue of which the conjunction of deontologism and the Bridge Principle obtains, there may well be more fundamental facts by virtue of which these facts obtain, and so on, right down to the most fundamental, bedrock facts of all. I am not necessarily after an explanation that is *bedrock*, in this sense. For example, in Chapter 4, I shall argue that the Bridge Principle obtains by virtue of certain connections between blameworthiness, control, and the nature of persons and consciousness. I do not think this explanation is shallow. But I shall not attempt to show that it is absolutely bedrock.

I should mention: in this dissertation, my strategy for helping us see how deontologism could be true never takes the form of offering a substantive argument for it. Rather, I shall develop a deontological picture of rationality, using as a goad for its development certain objections thought to be devastating. Once this picture is on the table, we shall be in a position to look and see that 'rational' and 'irrational' appear, at least in some contexts, to be used in the manner the picture predicts. As I said above, I am very open to the thought that there are multiple kinds of epistemic rationality. My hope is that my procedure will give us a firm grip on a *kind* of epistemic rationality of which deontologism turns out to be true, or at least plausible.

§1.3   Some Problems

In this section, I offer initial statements of the problems that will shape the development of the deontological explanation I shall offer. I begin with the problem which plays the most significant role in this dissertation. I then turn to three other problems.

§1.3.1 The At the Mercy of Evidence Objection

Deontologism says that believing *p* is rational for S if and only if believing *p* is blameless for S – that is, if and only if, in believing *p*, S fulfills epistemic obligations that S has at the time. In the literature, it is easy to find the objection that we do not have the ability to control our beliefs we would need to have, for it to be the case that we have epistemic obligations, in the first place. William Alston's memorable article on this topic is often cited.[20] So, for an initial characterization of this objection to deontologism, let us turn to Alston.

William Alston begins by arguing that we are unable to straightaway execute intentions to believe, or to not believe, particular propositions.[21] For example, I cannot straightaway execute an intention to believe that "the United States is still a colony of Great Britain", even with great monetary incentive.[22] Nor can I straightaway execute an intention to cease to believe deliverances of perception.

One might suggest that we can execute these kinds of intentions after a course of time, perhaps interrupted by other activities.[23] But Alston argues that whatever control we have of this kind is unreliable. Suppose that I have an intention to believe *p*. Suppose also that I search for evidence regarding *p*, find conclusive evidence in favor, and thereby believe *p*. Nevertheless, I did not believe *p by* executing my intention to believe *p*. Instead, the evidence I discovered caused my belief. For my believing to be the execution of my intention, it must be the intention, not the evidence, which causes my belief. But for this to be the case, I would need to succeed in "fighting very strong tendencies to believe when and only when something seems true".[24]

---

[20]   Alston 1989. See also Bergmann 2006, 77-78.
[21]   In Alston's (1989, 119-121) terminology, this is to say that we lack *direct voluntary control* over belief.
[22]   Ibid., 122. Plantinga (1993, 24) makes the same point.
[23]   This is the suggestion that we have *long-range voluntary control*, even if not direct voluntary control (Alston 1989, 133-134).
[24]   Ibid., 135.

Unsurprisingly, I will only occasionally succeed, if ever. So, even if we can occasionally execute intentions to believe over a course of time, this kind of control could only ever be unreliable.

Finally, one might suggest that we can *influence* what beliefs we have by executing intentions to actualize states of affairs that will cause us to have different beliefs than we would have otherwise had.[25] Alston grants that we have influence over some of our beliefs. For example, one can execute intentions to pursue evidence with care, courage, etc. – to be *intellectually virtuous*. And intellectual virtue would, in the long run, make some difference to which beliefs one holds.

We might even have an obligation to be intellectually virtuous. But, argues Alston, no such obligation could be *epistemic*.[26] For an obligation is epistemic only if fulfilling it places one in a "good position to get a true belief."[27] And intellectual virtue may not be enough to escape a poor position to get true beliefs. For example, suppose "S has lived all his life in an isolated primitive community where everyone unhesitatingly accepts the traditions of the tribe as authoritative."[28] Our hapless tribesman S might be admirably intellectually virtuous, yet believe traditions that are not truth-conducive. But then an obligation to be intellectually virtuous is not an epistemic obligation, because it is possible to fulfill it without achieving a "good position to get a true belief."

Put summarily, here is Alston's argument. Since our beliefs appear to be at the mercy of our evidence, we are not able to execute intentions to believe, whether straightaway or over a course of time. So, we have no obligations to exercise such control over belief. On the other

---

[25]   This is what Alston (Ibid., 137) calls *indirect voluntary influence*.
[26]   Ibid., 142-152.
[27]   Ibid., 144.
[28]   Ibid., 145.

hand, no obligation to merely influence our beliefs, perhaps by being intellectually virtuous, is an epistemic obligation. So, we do not have epistemic obligations. Therefore, deontologism is false.

Since a driving element of the objection is that belief appears to be at the mercy of evidence, I shall call this objection *the At the Mercy of Evidence Objection*. I have given merely an initial statement of it. I discuss it in more detail in Chapter 7.

§1.3.2 The Dilemma

Another problem facing deontological explanations comes in the form of a dilemma. Before discussing this dilemma, it is first necessary to single out a weak kind of accessibilism that I shall not discuss in this dissertation. On one model Williamson developed as a kind of concession to internalists, believing *p* is rational just in case *p* is among "what one does not know one does not know".[29] Now, it might be that whether *p* is among what one does not know one does not know is determined by the facts about which things are accessible to one.[30] If so, then Williamson's model vindicates accessibilism. But on this model, for believing *p* to be rational, it is enough that *nothing* is accessible to one, for then it follows that *p* is among what one does not know one does not know. In any case: on a weak version of accessibilism like this, rational believing does not require the accessibility of *anything*.

To lay some of my cards on the table, on the picture developed in this dissertation, accessibilism is motivated by the view that blameworthiness and blamelessness have application in the first place only when one is conscious of reasons to which one can respond. On this picture, the accessibilist theory finding explanation involves, as a built-in component, the view that rational believing requires the accessibility of reasons. So, I have a special interest in

---

[29]   Williamson 2013, 91-95. The quoted phrase occurs in a sentence merely defining what one believes; however, on this particular model, all belief is *ipso facto* rational.

[30]   In talking of what is accessible as 'the facts about which things are accessible', I disambiguate 'what is accessible' in a manner I discuss in Chapter 2.

accessibilist pictures that are stronger than Williamson's model, precisely because they involve this component.

In any case, let us focus on accessibilist pictures on which rational believing requires the accessibility of, at least, *something*. These accessibilist pictures face a dilemma. According to this dilemma, the motivation for the accessibilist picture either forces us to abandon foundationalism, or else loses its integrity.

Laurence BonJour and Michael Bergmann have offered this kind of dilemma.[31] It follows a recipe. First, we identify what it is about accessibility that explains why rational believing should require the accessibility of reasons. Perhaps their accessibility affects whether it is responsible to believe, or whether one has some level of assurance that forming belief would be correct.

In the next step of the recipe for the dilemma, we notice that, depending on how we understand accessibility, it is either true or false that the accessibility of a reason involves a *belief* regarding it. Take these possibilities one at a time. If accessibility involves belief, then perhaps it makes sense why a reason's accessibility affects responsibility or assurance. For what we believe affects what it is responsible for us to do, as well as whether we have assurance. But then the overall accessibilist picture stands in conflict with foundationalism, for it requires that, whenever believing is rational, one has other beliefs – the very kind of states also evaluable for rationality.

On the other hand, suppose accessibility does not involve belief. Then there is no conflict with foundationalism. However, given the etiolated nature of accessibility, it becomes unclear why the accessibility of certain reasons should be required for rational believing, in the first place. For how does what we do not so much as believe affect responsibility, or our level of

---

[31]   BonJour 1985, 69; Bergmann 2006, 11-13.

assurance? This is unclear. So, on this horn of the dilemma, the motivation for the accessibilist picture loses its integrity.

Let us call this objection *the Dilemma*. The Dilemma is a structural objection to attempts to motivate accessibilism, including deontological attempts. The precise flavor of the objection depends on the kind of motivation for accessibilism on offer. But the basic problem is that it seems hard to simultaneously preserve foundationalism and the integrity of the motivation for accessibilism. I shall give a fuller statement of the Dilemma when we return to it in Chapter 10.

§1.3.3 The Regress Objection to Accessibilism

Some have argued that accessibilism generates vicious regress.[32] According to this objection, accessibilism absurdly implies that an infinite hierarchy of sets of facts, each more complex than the last, must be accessible to the subject if she is ever to be rational. If this is right, then even if deontologism provides some traction for endorsing accessibilism, accessibilism itself only leads to disaster. And this shows that something has gone wrong. Let us call this *the regress objection*.

I have only given a rough statement of the regress objection to accessibilism. But there is no need to give more detail at this point. I address the regress objection in Chapter 2. As we shall see, I shall argue that accessibilism is *ambiguous*. Moreover, I shall argue that regress threatens on only one of its disambiguations. In addition, I shall argue that the motivations on offer for accessibilism only appear to support the regress-inducing disambiguation, but provide genuine support for the regress-free disambiguation. In this way, the regress objection offers us an opportunity to more clearly define accessibilism.

---

[32]   Wedgwood 2002a, 350-352; Bergmann 2006, 9-11, footnote 13; and, for relevant discussion, see Fumerton 1995, 81.

§1.3.4 Pessimism about Deontological Explanations

Now, Carl Ginet has tried to run a deontological explanation of accessibilism.[33] However, his explanation was found to fail, and for more than one reason. This failure gives the impression that these kinds of explanations are wrong-headed – unless and until we explain why features unique to Ginet's explanation, but not essential to deontological explanations in general, gave rise to it. The problem here is pessimism about deontological explanations.

I address pessimism about deontological explanations in Chapter 3. As we shall see, the failures of Ginet's explanation serve as opportunities to learn what deontologism would have to look like, if it is to explain accessibilism.

§1.4   Conclusion: The Strategy for Addressing the Four Problems

I have clarified what a deontological explanation is supposed to be, in the first place. I have also discussed, in a preliminary way, the main problems for deontological explanations which I shall address in this dissertation. These are the At the Mercy of Evidence Objection, the Dilemma, the regress objection, and pessimism about deontological explanations. In this concluding section, let me explain the method behind the sequence in which I shall address these four problems in coming chapters.

As I mentioned, replying to the regress objection generates a lesson about how to understand accessibilism. Similarly, dealing with pessimism about deontological explanations yields lessons about how to understand deontologism. In this way, solving these problems give us a better understanding of what a deontological explanation of accessibilism should look like, structurally speaking. For one, it clarifies exactly what the Bridge Principle should look like. So, it is a good place to start. I address the regress objection in Chapter 2, and pessimism about

---

[33]   Ginet 1975, 36.

deontological explanations in Chapter 3. This paves the way for offering a substantive

deontological explanation of accessibilism.

The next of these four problems which I shall address is the At the Mercy of Evidence

Objection. However, to effectively address this objection, I shall first need to put on the table a

substantive deontological explanation of accessibilism. Let me explain. According to the At the

Mercy of Evidence Objection, deontologism is false because we do not have epistemic

obligations, given that we lack control over belief. Whether this objection works depends on

what kind of control is presupposed by epistemic obligations. But if we are to use deontologism

to explain accessibilism, we need an understanding of epistemic obligations on which our having

these obligations would explain why epistemic rationality is determined by what is accessible.

So, to address the At the Mercy of Evidence Objection, we *first* need an account of epistemic

obligations that would allow us to explain accessibilism. *Then* we would need to show that the

kind of control presupposed by epistemic obligations, so understood, is a kind of control we have

over belief. The objection is answered only if we do both. And incidentally, things are similar

with the Dilemma. According to the Dilemma, the deontological motivation for the accessibilist

picture retains its integrity only if it sacrifices foundationalism. To see whether this Dilemma is

ultimately fatal, we need a concrete account of what that motivation *is*, so we can tell what it

takes for it to retain its integrity.

So, after dealing with the regress objection and pessimism about deontological

explanations in Chapters 2 and 3, I develop my substantive deontological explanation of

accessibilism in Chapters 4 and 5. This explanation in Chapters 4 and 5 gives rise to an

independent worry, which I deal with in Chapter 6. Then, in Chapters 7-9, I systematically

develop a reply to the At the Mercy of Evidence Objection. As we shall see, if this reply to the

At the Mercy of Evidence Objection is on track, it becomes fairly easy to reply to the Dilemma. I show this in the concluding chapter, Chapter 10, in which I dissolve the Dilemma.

Now that we have a sense of the method behind the sequence of the coming chapters, let us turn to the regress objection to accessibilism.

Chapter 2: Accessibilism Defined

§2.1   Introduction

Recall the gloss on accessibilism I gave in §1.2.2:

accessibilism   Whether believing *p* is rational for S is determined by what is accessible to S.

To recall from §1.3.3, some have argued that accessibilism generates vicious regress. According to this objection, accessibilism absurdly implies that an infinite hierarchy of sets of facts, each more complex than the last, must be accessible to the subject if she is ever to be rational. If this is right, then even if deontologism provides some traction for endorsing accessibilism, accessibilism itself only leads to disaster. And this shows that something has gone wrong. This is what I called *the regress objection*.

In this chapter, I shall argue that the regress objection rests on a misunderstanding of accessibilism. In particular, I shall argue that the phrase 'what is accessible to the subject' is ambiguous. This phrase may either refer to the *very things* accessible to the subject, or instead to the *facts about* which things are accessible to her. As we shall see, the regress objection only threatens the 'very things' disambiguation. Moreover, as I shall argue, it is only the 'facts about' disambiguation which enjoys genuine support from the motivations for accessibilism on offer. For these reasons, I shall recommend that accessibilism be disambiguated in the 'facts about' manner. The regress objection does not threaten accessibilism, so understood.

The chapter is structured as follows. In §2.2, I relate the regress objection. In §2.3, I establish that accessibilism is ambiguous in the manner stated above, and that the regress objection only threatens the 'very things' disambiguation. In §§2.4-2.5, I discuss the relationships between the motivations for accessibilism and the 'very things' and 'facts about' disambiguations. I conclude in §2.6.

§2.2   The Regress Objection to Accessibilism

Several have argued accessibilism generates vicious regress. I will focus on Ralph

Wedgwood's argument, which is the most developed.[34]

Remember that on our gloss of accessibilism, whether believing p is rational for S is

determined by what is accessible to S. Let me clarify the notion of determination in play here.

What determines whether something is the case is a set of facts. To say that whether something is

the case is determined by a set of facts of a particular kind is to say that, necessarily, if it is the

case (not the case), then a set of facts of this kind entails it is the case (not the case).[35] Following

Jim Pryor,[36] Wedgwood understands accessibilism as the view that whether believing p is

rational for S is determined by a set of facts with the following distinction: the set is accessible to

S. That is, Wedgwood understands accessibilism to be the following view:

(*)   Necessarily, if believing p is rational for S (not rational for S), then there is some
set of facts F such that F entails believing p is rational for S (not rational for S)
and F is accessible to S.

Wedgwood assumes the most common, standard understanding of accessibility,

according to which x is accessible to S just in case S is in a position to know x by reflection

alone. In this chapter, we shall go along with this standard understanding of accessibility. As we

shall see more clearly in §2.3, the ambiguity I shall highlight is not a matter of what relation 'is

accessible to' refers to, but instead a matter of what the phrase 'what is accessible to the subject'

refers to. As we will see in §2.5.3, knowability by reflection alone is not the only understanding

of accessibility.

---

[34]   Wedgwood 2002a, 350-352. Bergmann (2006, 9-10, footnote 13) presents a similar objection. For relevant discussion, see Fumerton
(1995, 81).
[35]   Thus, it is to say that whether it is the case *strongly supervenes* on a set of facts of this kind. See Van Cleve (1990, 225-226) and Kim
(2002, xvii) for definitions of strong supervenience.
[36]   Pryor 2001, 104.

The main premise of Wedgwood's regress objection is a view for which Timothy Williamson has influentially argued,[37] namely

> Anti-Luminosity   For all ordinary sets of facts F and persons S, F does not entail S is in a position to know F by reflection alone.

Succinctly, Williamson's argument is that for S to be in a position to know F by reflection alone, S could not be easily mistaken about F. But if F is an ordinary fact, it is possible for F to obtain in a case where S could easily be mistaken about F. Anti-Luminosity follows. Of course, Williamson's argument has been challenged.[38] What to make of these challenges is not a question I have the space to address, though.[39] But it is worth exploring whether accessibilists could accept Williamson's argument. So, for our purposes, we shall assume Anti-Luminosity.[40]

Anti-Luminosity is restricted to facts I have called *ordinary*. Williamson mentions that some extraordinary facts escape his argument.[41] Consider, for example, the fact that S exists. As there is no possible case in which S falsely believes S exists, there is no possible case in which S could be easily mistaken about it. So, Anti-Luminosity does not apply. If Williamson is right, though, Anti-Luminosity applies to the lion's share of facts "with which we engage in our everyday life", like the fact that S is cold, or that S is in pain.[42]

An accessibilist should want to allow not only that believing that S exists can be rational for S, but also that believing that S is in pain, for example, can be rational for S. At first glance, this may not seem like a problem for the accessibilist. For suppose S is in pain. Typically, this fact is accessible to S. So, can't the accessibilist hold that the fact that S is in pain entails that

---

[37]   Williamson 2000, Ch. 4.

[38]   See, e.g., Berker (2008) and Greco (2014, 194-195, especially footnote 50).

[39]   See Srinivasan (2015) for a defense of Williamson.

[40]   In his response to Williamson, Berker (2008) defends the view that the fact that it is rational to believe *p* (not rational to believe *p*) is a luminous fact. This view qualifies as a form of (*). For, on it, among the very things accessible to S is a fact entailing rationality status. Berker also says that "most epistemological internalists are…committed to luminosity claims of one form or another" (Ibid., 3, footnote 2). For all I have said, Berker's response to Williamson may be right. And he may also be right that many accessibilist-minded internalists endorse luminosity claims. But, even so, it is worth exploring the prospects of accessibilism, given Anti-Luminosity.

[41]   Williamson has two terms for the facts I am calling *extraordinary*: "trivial" and "curiosities" (Ibid., 108-109). I do not call them trivial because, as we shall see, some are quite substantive.

[42]   Ibid., 109.

believing that S is in pain is rational for S? If Wedgwood is right, she cannot coherently maintain this.

Here is the reasoning. Let 'Pain' refer to the fact that S is in pain, let p be the proposition that S is in pain, and let 'A( )' indicate that whatever is in the brackets is accessible to S. The following world is possible:

w_1 Pain & A(Pain)

But since Pain is an ordinary fact, Anti-Luminosity implies that w_2 is also possible:

w_2 Pain & ~A(Pain)

Now, let 'Other Facts' refer to other facts besides Pain which are candidates for entailing that believing p is rational for S. The following world is possible:

w_3 ~A(Other Facts)

But if a world is possible in which Pain & ~A(Pain) and a world is possible in which ~A(Other Facts), then a world is possible in which *both* of these conditions hold.[43] That is, w_4 is possible:

w_4 Pain & ~A(Pain) & ~A(Other Facts)

But (*) says any world in which believing p is rational for S is a world in which some set of facts entailing believing p is rational for S is accessible to S. In w_4, no set of facts entailing believing p is rational for S is accessible to S. So, on (*), it is not the case that believing p is rational for S in w_4. Therefore, on (*), Pain does *not* entail believing p is rational for S. For Pain obtains in a world in which it is not the case that believing p is rational for S, namely w_4.

Part of the trouble is that Pain is not accessible to S in w_2. So, the accessibilist might say that it is not Pain which is the fact which entails believing p is rational for S and is accessible to S in w_1, but instead A(Pain). In one way, this suggestion is of little help. For Anti-Luminosity

---

[43]   This step of the regress objection assumes that something along the lines of Lewis' (1986, 87-88) principle of recombination holds, according to which "patching together parts of different possible worlds yields another possible world".

implies A(Pain) & ~A(A(Pain)) is possible, and the reasoning runs as before to show that, on (*),

A(Pain) does not entail believing *p* is rational for S. But in another way, the suggestion indicates

a way out for the accessibilist. She can infinitely iterate accessibility. Where '…' signifies

infinitely many iterations, consider a fact we can refer to as follows:

$$A(A(A…(Pain)))…$$

Let 'Pain*' refer to the above fact. Just as infinity plus one is identical with infinity, A(Pain*) is

identical with Pain*. Thus, Pain* & ~A(Pain*) is not possible. So, Pain* is an extraordinary fact

over which Anti-Luminosity does not range. The accessibilist can hold that Pain* entails

believing *p* is rational for S.

Indeed, given the above reasoning, if (*) is true, Pain is a (proper or improper) part of a

fact that entails believing *p* is rational for S only if Pain is a part of Pain*. A similar conclusion

holds for other ordinary facts. An infinite regress of accessibility is always required. Some

philosophers accept this consequence of (*).[44] But, plausibly, this regress is vicious. Pain* is an

infinite array of facts each more complex than the last. Plausibly, it is never the case that every

member of an array of this kind is accessible to a mere mortal. So, if S is a mere mortal, it is not

the case that believing S is in pain, or cold, or etc., is rational for S. It may be that believing S

exists is rational for S, as the fact that S exists is one of the extraordinary few which escapes the

clutches of Anti-Luminosity. But even so, accessibilism is in bad straits.

In response, the accessibilist might point out that Anti-Luminosity is defined in terms of

knowability by reflection alone, and suggest that accessibility could be construed differently.[45]

But it is unclear how this suggestion could help (*). For (*) requires that a set of facts that entails

believing *p* is rational (not rational) for S is accessible to S. On this picture, in the crucial

---

[44]   See, e.g., Fales (2014).
[45]   I discuss another way of understanding accessibility in 2.5.3.

instance, accessibility must be a relation to *facts*. It is, at least, unclear what such a relation could be if it is not a kind of knowability. And Williamson's argument for Anti-Luminosity generalizes to any kind of knowability, so long as the relevant facts are ordinary. Though, for all I have said, a response to the regress objection on behalf of (*) might yet be in the offing, it is clear that the objection *threatens* (*), at the least. And this means that, insofar as we have reason to think accessibilism is (*), the objection threatens accessibilism.

§2.3   The Ambiguity of 'What is Accessible to the Subject'

In this section, I showcase an ambiguity in

accessibilism   Whether believing *p* is rational for S is determined by what is accessible to S.

and then show that the regress objection only threatens one of accessibilism's disambiguations. I begin by introducing the general kind of ambiguity I have in mind.

Consider

(a) Whether Abby is ready for a history exam is determined by what Abby knows.

(a) seems like a sensible thing to say. And it appears to have the following meaning: whether Abby is ready for a history exam is determined by the answer to the question of what Abby knows. The answer to the question of what Abby knows is a list of facts of the form *that Abby knows (does not know) p*. These facts are *facts about* which things Abby knows. (a) says that, necessarily, these facts either entail that Abby is ready for a history exam, or they entail that Abby is not ready.

It is possible, though uncharitable, to read (a) as saying that whether Abby is ready for a history exam is determined by the *very things* Abby knows. So read, (a) is demonstrably false. For nearly any set of facts F, there is a possible world in which Abby knows F, as well as a possible world in which F obtains but Abby does not know F. Abby might well be ready for a

history exam in the former world while not ready in the latter, even though F obtains in both worlds. Thus F neither entails that Abby is ready, nor entails that Abby is not ready.

Now consider

(b) Whether the star will go supernova is determined by what Abby the astrophysicist believes.

Read charitably, (b) says that among the very things Abby the astrophysicist believes are astrophysical facts that determine whether the star will go supernova. Perhaps Abby believes that the star meets conditions sufficient for its going supernova, or that the star meets conditions sufficient for its not going supernova. So understood, (b) says that, necessarily, these astrophysical facts either entail that the star will go supernova, or they entail that it will not.

It is possible, though uncharitable, to read (b) as saying that whether the star will go supernova is determined by the answer to the question of what Abby the astrophysicist believes. The answer to this question is a list of facts about which things Abby believes. So read, (b) is demonstrably false. Consider two possible worlds in which these facts are the same, where one of these facts is that Abby believes that the star meets conditions sufficient for its going supernova. But suppose that Abby believes this in one of the worlds simply because she unknowingly made a mistake in her calculations. It might well be that the star will go supernova in one world but not the other, contra (b) so construed.

Understood charitably, (a) has it that 'what Abby knows' refers to the answer to the question of what Abby knows, which is a set of facts about which things Abby knows. Meanwhile, understood charitably, (b) has it that 'what Abby the astrophysicist believes' refers not to the answer to the question of what Abby believes, but instead to the very things she believes, which include certain astrophysical facts. In this way, this kind of 'what' phrase may refer to the facts about which things the subject bears the attitude towards, or instead to the very

things towards which she bears the attitude. Some mechanism, like context, speaker intent, or etc., triggers one reading over the other.

Sometimes it is unclear which reading is at issue, as in

(c) Whether Abby will get the job is determined by what Abby knows.

Interestingly, different intonation patterns offer a rough and ready way to indicate which reading of (c) is intended. If one says that *what* Abby knows determines whether she will get the job, we take her to mean that among the very things Abby knows are facts that determine the matter. But if one says that what Abby *knows* determines whether she will get the job, we take her to mean that the facts about which things Abby knows determine the matter.

Now we have a handle on the kind of ambiguity that infects accessibilism. We should expect that 'what is accessible to S' is ambiguous in the same way that phrases like 'what Abby believes' and 'what Abby knows' are ambiguous. Therefore, we should expect that

> *accessibilism* Whether believing p is rational for S is determined by what is accessible to S.

is ambiguous between $(A)^{\text{very things}}$ and $(A)^{\text{facts about}}$:

> $(A)^{\text{very things}}$ Whether believing p is rational for S is determined by the *very things* accessible to S.

> $(A)^{\text{facts about}}$ Whether believing p is rational for S is determined by the *facts about* which things are accessible to S.

On the standard understanding of accessibility, the very things accessible to S are the things S is in a position to know by reflection alone. Such things are, presumably, facts. And if the very things accessible to S are facts, the ambiguity in the reference of 'what is accessible to S' is between the *first-order* facts which are accessible to S, and the *second-order* facts about which first-order facts are accessible to S.[46]

---

[46] I thank an anonymous referee for recommending this additional way of framing the disambiguation.

Notice that Audi's definition of accessibilism, quoted in §1.2.2, can be read as either

(A)$^{\text{very things}}$ or (A)$^{\text{facts about}}$. The following bracketed insertions clarify the different readings:

> …justification is grounded entirely in [*the very things which are*] what is internal to the mind, in a sense [of 'internal'] implying that [something internal is such that] it is accessible to introspection or reflection by the subject…

> …justification is grounded entirely in [*the answer to the question of*] what is internal to the mind, in a sense [of 'internal'] implying that [something internal is such that] it is accessible to introspection or reflection by the subject…

The first reading is (A)$^{\text{very things}}$: rationality is determined by the very things accessible to S. The

second reading is (A)$^{\text{facts about}}$: rationality is determined by the answer to the question of what is

accessible to S. The answer to this question is the facts about which things are accessible to S.

We have seen that accessibilism is ambiguous between (A)$^{\text{very things}}$ and (A)$^{\text{facts about}}$. And

now remember the regress objection to accessibilism. According to that objection, accessibilism

is to be understood as

> (*)   Necessarily, if believing *p* is rational for S (not rational for S), then there is some set of facts F such that F entails believing *p* is rational for S (not rational for S) *and F is accessible to S.*

But when (*) is combined with Anti-Luminosity, it became inevitable that believing *p* is rational

for S only if each member of an infinite array of facts, each more complex than the last, is

accessible to S. And that is absurd.

This regress objection only threatens (A)$^{\text{very things}}$. (A)$^{\text{very things}}$ is equivalent to (*). On

(A)$^{\text{very things}}$, some set of facts entailing S's rationality status must *itself* be accessible to S. And

that is what (*) says, too. But (A)$^{\text{facts about}}$, on the other hand, merely says that some set of facts

entailing S's rationality status must be among *the facts about* which things are accessible to S.

These facts about which things are accessible to S need not *themselves* be accessible to S. Thus,

one who simply endorses (A)$^{\text{facts about}}$ can coherently hold that A(Pain) entails believing *p* is

rational for S, even if

w_5 A(Pain) & ~A(A(Pain)) & ~Other Facts

is possible. For (A)$^{\text{facts about}}$ does not require A(Pain) to be accessible in order for believing p to

be rational for S. Thus, A(Pain) can entail that believing p is rational for S, all by itself.[47] There

is no foothold for regress in (A)$^{\text{facts about}}$.

So, whether the regress objection hits its mark depends on the 'very things/facts about'

ambiguity in (A). (A)$^{\text{facts about}}$ emerges entirely unscathed. (A)$^{\text{very things}}$ does not.

§2.4   Two Understandings of Accessibilism's Motivations

From the fact that the regress objection threatens (A)$^{\text{very things}}$, but not (A)$^{\text{facts about}}$, it does

not follow that accessibilism is out of the woods. For what if, for all that, (A)$^{\text{very things}}$ is the

proper disambiguation of accessibilism? The way forward is to discern which of (A)$^{\text{very things}}$ and

(A)$^{\text{facts about}}$ is the genuine upshot of the intuitive motivations by virtue of which accessibilism

exists, in the first place. I reviewed these motivations in §1.2.2. They are: BonJour's example of

Norman the clairvoyant, the new evil demon problem, and deontologism. If (A)$^{\text{facts about}}$ not only

escapes the regress objection unscathed, but is also the genuine upshot of these motivations, I

suspect it is fair to conclude that (A)$^{\text{facts about}}$ better captures the heart of accessibilism.

In this section, I will show that the motivations for accessibilism on offer appear to

support *both* (A)$^{\text{very things}}$ and (A)$^{\text{facts about}}$. In the next section, though, I will argue that this

appearance actually depends on a mistake, and that only (A)$^{\text{facts about}}$ enjoys genuine support.

That these motivations *appear* to support (A)$^{\text{very things}}$ can serve as an error theory

explaining why some may have mistaken (A)$^{\text{very things}}$ for the heart of accessibilism. And an error

theory appears to be needed. We have seen that Audi's definition of accessibilism can be read as

---

[47]   And so, without the help of Other Facts, as w_5 makes clear.

either $(A)^{\text{very things}}$ or $(A)^{\text{facts about}}$. But it is easy to find definitions of accessibilism which can only

be read as $(A)^{\text{very things}}$. For example, Jim Pryor defines accessibilism as the view that

> Whether one is justified in believing *p* supervenes on facts which one is in a
> position to know about by reflection alone.[48]

While 'what is accessible' can be read as referring to the answer to the question of what is

accessible rather than the very things accessible, "facts which one is in a position to know about

by reflection alone" can only be read as referring to the very things accessible to one. A similar

point holds with respect to Carl Ginet's view according to which

> Every one of every set of facts about S's position that minimally suffices to make
> S, at a given time, justified in being confident that p must be *directly recognizable*
> to S at that time.[49]

So, we need an error theory explaining why some take $(A)^{\text{very things}}$ to be accessibilism. It is also

clear that if the heart of accessibilism has been misunderstood, what will show this is a better

understanding of the intuitive data which motivates the view in the first place, not a head count

of definitions in terms of how they can be read.

Let us begin with how the motivations for accessibilism appear to support $(A)^{\text{very things}}$.

Return to the case of Norman the clairvoyant. BonJour allows that the fact that Norman is

reliable is, in some sense, a *reason* to believe that the President is in NYC. In BonJour's words,

this fact is among the "true premises or reasons…that could in principle provide a basis for

justification".[50] But BonJour does not allow that this fact could be a reason explaining why it is

rational for *Norman*, by contrast with a third party to whom the fact is accessible, to believe the

President is in NYC.

[48]   Pryor 2001, 104.
[49]   Ginet 1975, 34.
[50]   BonJour 1985, 43.

Much of what BonJour says suggests that, in the following way, deontologism undergirds the idea that a reason can help explain why a person is rational only if it is accessible to her.[51] Given deontologism, rationality consists in its being blameless for one to believe. But whether it is blameworthy to believe depends solely on reasons that can *guide* one – that is, reasons one can take into account in one's reasoning[52] and either believe or not believe on their basis. For a reason which cannot guide one in this sense is irrelevant to whether or not it is blameworthy for one to believe. But a reason can guide one only if it is accessible to one. Therefore, connecting the dots, only reasons accessible to one can help explain why one is rational.

We can represent the argument as follows:

(1) Rationality consists in its being blameless for S to believe.

(2) If (1), then only reasons which can guide S can help explain why S is rational.

(3) Only accessible reasons can guide S.

(4) So, only accessible reasons can help explain why S is rational. [from (1), (2), and (3)]

Among the facts which entail S is rational are some which explain why S is rational. According to (4), all of the latter are accessible to S. And that implies (A)$^{\text{very things}}$, on which some set of facts which entails S is rational is accessible to S, if S is rational.

(1)-(4) would explain the intuition that Norman's reliability could not make him rational. It is not accessible to him that he is reliable. Thus, he cannot be guided by that fact, and so it cannot help explain why he should be rational. Similarly, (1)-(4) could explain the intuition that the unreliability of the people in the demon world could not make them irrational. Their unreliability is not accessible to them. Thus, they cannot be guided by the fact that they are

---

51   BonJour Ibid., 42-43.
52   I am liberal with respect to what qualifies as reasoning. In addition to deliberation and inference, I also take non-inferential belief formation, revision, and abandonment to be events of reasoning. See Wedgwood (2006, 660-661) for precedence in this regard.

unreliable, and so it cannot help explain why they should be irrational. And notice that because (1) is deontologism, each of the motivations for accessibilism on offer nicely dovetail in (1)-(4) – namely, the example of Norman, the new evil demon problem, and deontologism.

I shall later argue that (1)-(4) rests on a mistake. If that is right, then the initial attractiveness of (1)-(4) gives us an error theory which would explain why some may have mistaken (A)$^{\text{very things}}$ for the heart of accessibilism. But before digging further into (1)-(4), let us see if the motivations for accessibilism also appear to support (A)$^{\text{facts about}}$.

Consider Norman again. BonJour gives us facts about which things are accessible to Norman. We are told, in particular, that no reason to believe the President is in NYC today is accessible to him. Intriguingly, we have the intuition that it is not rational for Norman to believe the President is in NYC *as soon as we are told these facts about which things are accessible to him*.

Why is this enough to generate the intuition? Let us begin again with deontologism, on which rationality consists in its being blameless for one to believe. Plausibly, the facts about which reasons can guide S determine the facts about whether it is blameless for S to believe. For suppose that the facts about which reasons can guide two different persons are the same. Then it is intuitive that, should they believe a particular proposition, it would not be appropriate to blame one unless it were also appropriate to blame the other. Now add to the mix the proposal that the facts about which things are accessible to S determine the facts about which reasons can guide S. This proposal is *prima facie* plausible. And the result, to connect the dots, is that the facts about which things are accessible to S determine the facts about whether S is rational.

We can summarize the argument as follows:

(5) Rationality consists in its being blameless for S to believe.

(6) If (5), then whether believing p is rational for S is determined by the facts about which reasons can guide S.

(7) The facts about which reasons can guide S are determined by the facts about which things are accessible to S.

(8) So, whether believing p is rational for S is determined by the facts about which things are accessible to S.   [from (5), (6), and (7)]

(8) is $(A)^{\text{facts about}}$.

(5)-(8) would explain why we have an intuition about whether Norman is rational as soon as we are told the facts about which things are accessible to him. For these facts determine the facts about which reasons can guide him, which facts, in turn, determine the facts about whether he is rational. Similarly, (5)-(8) may also explain why we have the intuition that people in the demon world are exactly as rational as those in the normal world. For, plausibly, the facts about which things are accessible are the same in each world. Again, as (5) is deontologism, each of the motivations for accessibilism on offer nicely dovetail with this (5)-(8) argument for $(A)^{\text{facts about}}$.

But we have seen that the same appears to be true of the (1)-(4) argument for $(A)^{\text{very things}}$. So we are left with a question: is $(A)^{\text{very things}}$ or instead $(A)^{\text{facts about}}$ the genuine upshot of these motivations?

§2.5   The Better Way to Understand Accessibilism's Motivations

In §2.3, we saw that whether the regress objection hits its mark depends on the 'very things/facts about' ambiguity. In this section, I shall argue that whether the motivations for accessibilism provide genuine support also depends on this ambiguity. In §2.4, we saw that the motivations for accessibilism *appear* to support both $(A)^{\text{very things}}$ and $(A)^{\text{facts about}}$. In the heart of the (1)-(4) argument for $(A)^{\text{very things}}$, though, is the idea that only reasons which can guide S can help explain why S is rational. I will show that this idea conflates two different kinds of reasons,

and thus that (1)-(4) is confused. Moreover, we will see that the (5)-(8) argument for (A)^facts about

is what the motivations for accessibilism genuinely support, once we are clear on the distinction

between these two kinds of reasons. And finally, though (5)-(8) employs substantive,

controversial premises, we shall see that it is defensible.

§2.5.1 Two Kinds of Reasons

(1)-(4) mentions reasons which can guide S as well as reasons which can help explain

why S is rational. It is entirely appropriate to use the word 'reasons' in each of these contexts.

However, we must be careful to not let this fact about our linguistic practice blind us to the

possibility that what can guide the subject may be very different from what explains why she is

rational. As it turns out, these are importantly different kinds of reasons.

What are reasons which can guide the subject? Suppose Beth has had the flu for the past

few days. This morning, she is feeling slightly better, and is wondering whether or not she is well

enough to go to work. Upon introspection, she knows she no longer feels nauseous. But she also

notices a headache. Now suppose also that, unbeknownst to her, her respiratory tissue is

inflamed. Some of the reasons in play in this case can guide Beth to either believe she is well

enough or instead refrain from believing this. But not all can. Notably, the fact that her

respiratory tissue is inflamed cannot guide her in this way. This is because, in a certain sense of

'could', that her respiratory tissue is inflamed is not something for which she could believe or

refrain – or, more precisely, it is not something that could move her some degree toward

believing or refraining. For she is entirely unaware that it is inflamed, and so this is not

something she can take into account in her reasoning. By way of contrast, she can take into

account the fact that she no longer feels nauseous, and the fact that she has a headache.

Let me clarify how I will be thinking about this example. To use terminology common in

the literature on practical reason but equally applicable to epistemology, let us call the reasons

for which or *on the basis of which* one did something one's *motivating reasons* for having done it.[53] Now, suppose it turns out that Beth's headache tips the balance in her deliberation about what to believe, and she believes she is not well enough to go to work. Depending on one's background theory about motivating reasons, her motivating reason for believing she is not well enough is either her *belief* that she has a headache, or simply *that she has a headache*.[54] The main argument of this paper does not depend on which way of thinking about motivating reasons we adopt. However, I myself find it more natural to hold that what Beth took into account in her reasoning was the *content* of her belief, not the belief itself. So, to fix ideas, henceforth I shall treat *that she has a headache* as her motivating reason, in the development of the example in which she goes on to believe she is not well enough.

Return to Beth at the time when she is still wondering whether she is well enough. A reason can guide her if and only if, in a certain sense of 'could', it *could* be a motivating reason for which she believes or instead refrains. Let us call any reason that could motivate, in this sense, a *potentially motivating reason*.[55] To say that among the reasons that can guide her is that she has a headache, but not that her respiratory tissue is inflamed, is to say that the first is a potentially motivating reason, while the second is not. Reasons which can guide the subject are one and the same as potentially motivating reasons.

It is a substantive question what sense of 'could' is at issue, when we say a potentially motivating reason *could* motivate. To illustrate, there are surely possible scenarios in which Beth learns of the inflamed tissue, and the fact that the tissue is inflamed motivates her to believe she is not well enough. But, presumably, the sense in which a potentially motivating reason could

---

[53]   For precedence, see Schroeder (2007, 12).
[54]   For the former kind of background theory, see Smith (1987); for the latter, see Schroeder (2008).
[55]   Bernard Williams (1995, 35) thinks of reasons as starting points for a "deliberative route" to a conclusion. Kieran Setiya (2007, 12) has a similar picture. In these terms, potentially motivating reasons are starting points in reasoning. Schroeder (2007, 14) calls potentially motivating reasons "subjective normative reasons".

motivate is more immediate: it is something one can take into account in one's reasoning at the relevant time.[56] I will return to this issue later.

I turn now to reasons which help explain why one is rational. Suppose we recognize it is rational for Beth to believe she is not well enough to go to work. Still, we might ask *what explains* the fact that she is rational. In answer to this question, we try to cite *reasons why* she is rational. Such reasons are part of an explanation of the fact that she is rational. Similarly, if we recognize Beth is not rational, we might ask what explains this fact. Then we would try to cite reasons that explain the fact that she is not rational. I will call this kind of reason a *rationality explanation reason*.[57]

We are now well situated to evaluate the idea I had said is in the heart of (1)-(4), which we can restate as the idea that only potentially motivating reasons are rationality explanation reasons. I suspect the idea has nothing to recommend it besides a failure to appreciate the difference between these two kinds of reasons.

§2.5.2 The Confusion Behind (1)-(4)

Let us rewrite (1)-(4) using the terminology of potentially motivating reasons and rationality explanation reasons:

(1) Rationality consists in its being blameless for the subject to believe.

(2) If (1), then only potentially motivating reasons are rationality explanation reasons.

(3) Only accessible reasons are potentially motivating reasons.

(4) So, only accessible reasons are rationality explanation reasons. [from (1), (2), and (3)]

---

[56]   The fact that her respiratory tissue is inflamed does appear to be, in *some* sense, a reason to believe that Beth is not well enough to go to work. This is similar to the sense in which the fact that Norman is reliable is a reason to believe that the President is in NYC. These facts *favor* believing, much as a proposition's brute *truth* counts in favor, in some sense, of believing it. Schroeder (Ibid., 13) calls facts like these "objective normative reasons".

[57]   Rationality explanation reasons are a subtype of the reasons Schroeder calls "explanatory" (Ibid., 11). Like John Broome (2013, 47-49), I understand explanation to be a relation between worldly things, i.e., facts, rather than a relation between linguistic descriptions.

(4), again, immediately implies (A)^{very things}. At least one premise of this argument is plausible, namely (3). But, as I hope to show, if (3) is true, (2) is false.

Assume (3) is true: only accessible reasons are potentially motivating reasons. Plausibly, it follows that for any given potentially motivating reason, the fact that it is accessible is at least part of what explains the fact that it is a potentially motivating reason. Now remember Beth, who is wondering whether she is well enough to go to work. Suppose it is rational for Beth to believe she is not well enough. Let us also suppose, as we are free to do, that part of what explains why Beth is rational to believe she is not well enough is *the fact that one of the potentially motivating reasons she has is that she has a headache*. From what we have said, at least part of what explains this last fact is *the fact that it is accessible to her that she has a headache*.

Thus, the fact that it is accessible to Beth that she has a headache is part of what explains why Beth is rational to believe she is not well enough. That is, this fact is a rationality explanation reason. But surely it is possible for the fact that it is accessible to Beth that she has a headache to *fail* to be a potentially motivating reason. That Beth has a headache is one thing, and the fact that it is accessible to her that she has a headache is another. We should expect that what it takes for the first to potentially motivate is different from what it takes for the second. For one, we should expect that what it takes for the first to be accessible is different from what it takes for the second. So, since the fact that it is accessible to her that she has a headache is a rationality explanation reason, it is simply not true that only potentially motivating reasons are rationality explanation reasons.

One way to avoid this result is to hold that if *p* is accessible to S, it follows by metaphysical necessity that it is accessible to S that *p* is accessible to S. An initial point is that, given the standard understanding of accessibility as knowability by reflection alone, this strategy commits one to denying Anti-Luminosity. But even if we understand accessibility differently, we

should expect the relevant analogue of Anti-Luminosity to be at least as plausible as Anti-Luminosity itself. No matter how we understand accessibility, the fact that Beth has a headache is one thing, and the fact that it is accessible to her that she has a headache is another. We should expect it to be possible for the first to obtain without the second.

In any case, I conclude that it is not true that only potentially motivating reasons are rationality explanation reasons. Put differently,

> (2) If rationality consists in its being blameless for the subject to believe, then only potentially motivating reasons are rationality explanation reasons.

has a false consequent. But (2)'s antecedent is just (1), i.e., deontologism. So either (1) or (2) is false. (1)-(4) is doomed.

But the moral we should draw, in particular, is that (2) is false. Surely a deontologist can hold that *the fact that* something is a potentially motivating reason may be a rationality explanation reason even if this fact is not itself a potentially motivating reason. Similarly, a deontologist can hold that the fact that something is accessible may be a rationality explanation reason even if this fact is not itself accessible. To put it colloquially, this is simply to hold that it may well be blameless for one to believe *p* partially in virtue of *the fact that* one has a particular guide, even if the fact that one has that guide is not *itself* a guide one has. This looks like something we should expect a deontologist to endorse, once clear on the issue. Our conclusion that not all rationality explanation reasons are potentially motivating reasons shows that (2) is false, not that deontologism is false.

So what makes (2) seem attractive? When I first introduced the (1)-(4) argument, the following idea motivated (2):

> Irrelevance   A reason which cannot guide S is irrelevant to whether or not it is blameworthy for S to believe.

For reasons I shall now explain, any attraction to the thought that Irrelevance is good grounds for (2) derives entirely from conflating potentially motivating and rationality explanation reasons.

It is natural to hold that potentially motivating reasons play an indispensable role in explanations of its being blameless for one to believe. So if 'a reason' is read as only ranging over candidates for being potentially motivating reasons, and so only over candidates for playing this indispensable role, Irrelevance appears plausible. Any such candidate which cannot guide the subject fails to be a potentially motivating reason, and so is disqualified from playing this role. But so construed, Irrelevance does not imply (2). For it may well be that an explanation of its being blameless for one to believe involves, besides certain potentially motivating reasons, the fact that one *has* those potentially motivating reasons, and whatever explains that fact.

On the other hand, suppose 'a reason' ranges over candidates for rationality explanation reasons as the deontologist understands them – that is, parts of explanations of its being blameless for one to believe. Then Irrelevance implies (2): if deontologism is true, only potentially motivating reasons are rationality explanation reasons. But so construed, Irrelevance is implausible. As we have seen, *the fact that* one has a particular guide may not *itself* be a guide she has, but it makes perfect sense to suppose that this fact may help explain why it is blameless for her to believe. These facts are far from irrelevant.

So, Irrelevance is plausible only if concerned with potentially motivating reasons, but implies (2) only if concerned with rationality explanation reasons. Hence, I suspect the temptation to endorse (2) derives from one's initially taking the reasons at issue to be potentially motivating reasons, but then failing to distinguish such reasons from rationality explanation reasons. To do this, though, is to conflate two very different kinds of reasons.

As we shall see, the (5)-(8) argument for (A)$^{\text{facts about}}$ is what the motivations for accessibilism genuinely support, once we are clear on the distinction between these two kinds of reasons.

§2.5.3 On (5)-(8)

Let us rewrite (5)-(8) in terms of potentially motivating reasons:

(5) Rationality consists in its being blameless for S to believe.

(6) If (5), then whether believing *p* is rational for S is determined by the facts about which potentially motivating reasons S has.

(7) The facts about which potentially motivating reasons S has are determined by the facts about which things are accessible to S.

(8) So, whether believing *p* is rational for S is determined by the facts about which things are accessible to S.   [from (5), (6), and (7)]

Look at (6). For the deontologist, rationality explanation reasons are facts which explain the facts about whether it is blameless for S to believe. Now notice that according to (6), it is the *facts about* which potentially motivating reasons one has, not these *very* reasons themselves, which determine the facts about whether it is blameless for S to believe, and so constitute rationality explanation reasons. Here the 'potentially motivating reason/rationality explanation reason' distinction is properly appreciated. And appreciating this distinction leads to a particular disambiguation of the 'very things/facts about' ambiguity, as can be seen. It is no accident that the conclusion of the above argument is (A)$^{\text{facts about}}$ and not (A)$^{\text{very things}}$.

Now, by contrast with (2), (6) has some plausibility. Plausibly, the facts about which potentially motivating reasons S has determine the facts about whether S has *adequate* potentially motivating reasons to believe. And plausibly, the facts about whether S has adequate potentially motivating reasons to believe determine the facts about whether it is blameless for S

to believe. It follows that the facts about which potentially motivating reasons S has determine the facts about whether it is blameless for S to believe, just as (6) says.

Let us consider the other premises of (5)-(8). (7) follows from two premises:

(7a) Only accessible things are potentially motivating reasons.

(7b) For all x, whether x is a potentially motivating reason S has is determined by the facts about what x is (intrinsically)[58] and whether x is accessible to S.

If (7a) is true, then the facts about which things are accessible to S determine the facts about which potentially motivating reasons S has so long as any two subjects have the same potentially motivating reasons if the very same set of items is accessible to each. Now suppose it were possible that, for some item x, in addition to the facts about what x is and whether x is accessible, we need to add other facts about S or the environment in order to fix whether x is a potentially motivating reason S has. Then two subjects can have different potentially motivating reasons even though the very same set of items is accessible to each. (7b) rules out precisely that possibility.

Let us begin with (7a). Recall that it is a substantive question what sense of 'could' is at issue, when we say potentially motivating reasons *could* motivate. Remember Beth, who can discern upon introspection that she has a headache. In the relevant sense of 'could', we said that her having a headache could motivate her to believe she is not well enough to go to work, while the fact that her respiratory tissue is inflamed, of which she has no inkling, could not motivate her. I suspect the sense of 'could' at issue is conceptually tied to the sense of 'can' at play in the notion of accessibility. Accessibility is access one *can* have, in a certain sense of 'can'. For example, if direct recognition, knowledge by reflection alone, or etc., is access, then direct

---

[58] What we want (7b) to rule out is the possibility of two subjects having different potentially motivating reasons in a case where the very same set of items is accessible to each. But if we allow part of *what x is* to consist in x's having extrinsic or relational properties, this possibility is not ruled out even if whether x is a potentially motivated reason is determined by the facts about what x is and whether x is accessible. For then, supposing x is accessible to two different subjects, a fact about what x is could involve extrinsic properties by virtue of which x potentially motivates one of these subjects but not the other.

recogniz*ability*, know*ability* by reflection alone, or etc., is access one can have, in this sense. My suspicion, in particular, is that x is *accessible* just in case one can access x in the strongest sense[59] in which one must be able to access x if x is to be a potentially motivating reason for one. If this is right, then, as a matter of course, all potentially motivating reasons are accessible.

Here is why I have this suspicion. Accessibility is a technical notion within accessibilism, a view often linked with deontologism. And deontologism encourages a particular way to think about potentially motivating reasons. If for it to be rational to believe is for it to be blameless to do so, then, in the relevant cases,[60] one's believing is a response to reasons it is appropriate to hold one responsible to. That is, it is a response to reasons it would be appropriate to blame one for responding to in one way rather than another. It is natural for deontologists to think of potentially motivating reasons as precisely the reasons it is appropriate to hold one responsible to, in this sense. But it is appropriate to hold one responsible to a given reason only if she is able to take it into account in her reasoning at the time, which requires that, in some fairly strong sense of 'can', she can access it at the time.

One possibility is that a person can take a reason into account in the relevant sense only if she is in a position to know it by reflection alone. If so, then deontologism vindicates the standard understanding of accessibility as knowability by reflection alone. Thus far in the paper, we have gone along with this standard understanding of accessibility. But for all I have argued, it is an open question whether being appropriately held responsible to a reason requires that the person be in a position to know the reason by reflection alone. Perhaps, instead, what is required

---

[59] One sense in which one must be able to access it is that there is some metaphysically possible world in which one accesses it. A significantly stronger sense is required.

[60] The deontologist may want to allow that some ways of coming to believe fall outside the scope of her view. These are ways that would be screened off from the question of whether it is blameless for one to believe, and so screened off from the question of whether one is rational to believe. An example of such a way of coming to believe might be Dr. Evil shooting his belief altering ray gun at one. Even if, in this sense, there is a way of coming to believe such that if I came to believe in that way, I could not be appropriately blamed, it does not follow that, *in the relevant sense*, it is blameless for me to believe.

is simply that the person must be *conscious* of the reasons, whether via reflection or in some other way, such as sense perception. Indeed, in Chapter 4, I shall simply argue that being appropriately held responsible to a reason requires being conscious of it.

In any case, given that x is a potentially motivating reason just in case it is appropriate to hold one responsible to x, and given that this is appropriate only if the person can take x into account in her reasoning, it is plausible that there is *some* sense in which one must be able to access x if x is to be a potentially motivating reason. For, plausibly, a person can take x into account in her reasoning only if she can access x, in some sense. Given that accessibility is access one can have, in whatever sense turns out to be required, the upshot is that only accessible things are potentially motivating reasons. This upshot is (7a).

Let us turn to

(7b) For all x, whether x is a potentially motivating reason S has is determined by the facts about what x is (intrinsically) and whether x is accessible to S.

One might object to (7b) in the following way. No matter what the item is that is accessible to one, it is always *possible* for one to have some kind of disability that prevents it from potentially motivating one.[61] But then, in addition to the facts about what x is and whether x is accessible, we need to add the fact that the subject has the relevant abilities, to fix whether x is a potentially motivating reason.

A response to this objection is available. We can endorse the view that part of what it *is* for a mental state to be an accessing of x, in particular, is for the one in this state to have a disposition to transition from this state to a range of other mental states, where this range of states is fixed by what x is.[62] With this view in hand, we could say that, for any item of the

---

[61]   Someone with this disability might be like the Tortoise in Carroll's (1895, 278-280) fable.
[62]   See, e.g., Levin (2013, §4.3) and Wedgwood (2007, 165-167).

relevant kind, to have a disability that prevents it from potentially motivating one is for it to *fail* to be accessible to one in the first place.

Moreover, if this response is on track, it follows that (7b) is true. For then any item of the relevant kind, *if* accessible in the first place, is a potentially motivating reason. And this is to say that the facts about what x is and whether x is accessible suffice to fix whether x is a potentially motivating reason: (7b) is true. Of course, the envisioned response to the objection involves a substantive, likely controversial view.[63] The point is merely that (7) is defensible.

The only remaining premise is (5), which affirms deontologism. What to make of deontologism is a major issue I do not address in this chapter. Some argue it is a mistake to think of rationality as blamelessness, in the first place.[64] Others argue that deontologism presupposes we have a kind of control over belief that we do not, in fact, possess.[65] By the end of the dissertation, when my deontological explanation of accessibilism is on the table, we will have a better grasp on these issues. This is why I wait until §10.6 before I suggest that deontologism is true.

In §2.4, I explained how (A)[very things] and (A)[facts about] each appears to be supported by the motivations for accessibilism, namely the case of Norman, the new evil demon problem, and deontologism. Each of the two ways of understanding these motivations appealed to the plausible idea that only accessible things are potentially motivating reasons. However, the (1)-(4) argument for (A)[very things] conflates potentially motivating reasons and rationality explanation reasons. Once we are clear on the distinction between these two kinds of reasons, it is the (5)-(8) argument for (A)[facts about] which is the real upshot of the plausible idea that only accessible things

---

[63]   Even if we shy away from the envisioned response, notice that the objection does not touch a view otherwise identical to (A)[facts about] save one revision. This is the view that whether believing *p* is rational for S is determined by the facts about which things are accessible to S *plus the facts about which abilities S has*. Even if (A)[facts about] turns out to be false, to offer a starting point from which to take, if needed, these kinds of small steps from disaster is a service to anyone who finds accessibilism's motivations attractive.
[64]   See, e.g., Wedgwood (2002a, 351) and Pryor (2001, 114-115).
[65]   See especially Alston (1989, 115-152).

are potentially motivating reasons. Finally, we have seen that while (5)-(8) employs substantive premises, the argument is defensible. So, it is (A)$^{\text{facts about}}$, not (A)$^{\text{very things}}$, which enjoys genuine support from the motivations for accessibilism.

§2.6   Conclusion

After relating the regress objection to accessibilism, I showed that accessibilism is ambiguous because 'what is accessible to the subject' is ambiguous. The phrase either refers to the very things accessible to the subject, or instead to the facts about which things are accessible to her. I showed that the regress objection threatens (A)$^{\text{very things}}$ but not (A)$^{\text{facts about}}$. After this, I explained why the motivations for accessibilism appear to support both (A)$^{\text{facts about}}$ and (A)$^{\text{very}}$ $^{\text{things}}$. This appearance serves as an error theory that would explain why some might have taken (A)$^{\text{very things}}$ to be the heart of accessibilism. But I argued that once we are clear on the distinction between potentially motivating reasons and rationality explanation reasons, it is (A)$^{\text{facts about}}$, not (A)$^{\text{very things}}$, which enjoys genuine support from the motivations for accessibilism. So, not only is it true that whether the regress objection hits its mark depends on the 'very things/facts about' ambiguity, whether the motivations for accessibilism provide genuine support depends on this ambiguity, as well. Compared to (A)$^{\text{very things}}$, (A)$^{\text{facts about}}$ gets the better end of the stick on both counts: it entirely escapes the objection, and enjoys genuine support from the motivations.

For these reasons, I submit that future discussions of accessibilism should focus on (A)$^{\text{facts about}}$. That is, we should define accessibilism as follows:

accessibilism   Whether believing *p* is rational for S is determined by the facts about which things are accessible to S.

I note that my substantive deontological explanation in Chapter 4 shall vindicate the (5)-(8) argument for accessibilism so understood. That is a nifty result, as (5)-(8) captures the

intuitive motivations for accessibilism on offer, including the example of Norman the clairvoyant and the new evil demon problem.

In this chapter, we have gained a clearer understanding of accessibilism, by getting our hands dirty with the regress objection. In the next chapter, we shall address pessimism about deontological explanations. The result will be a clearer understanding of deontologism, insofar as deontologism has any hope of explaining accessibilism.

Chapter 3: Is Deontologism Explanatorily Impotent?

§3.1   Introduction

Does deontologism explain accessibilism? If so, how? One way to get traction on these questions is to look at prior attempts to use deontologism to explain accessibilism. In this chapter, I shall consider a deontological explanation offered by Carl Ginet. As we shall see, Ginet's explanation fails, and in more than one respect. The cumulative effect of these failures, by my lights, is the impression that deontologism is explanatorily impotent. This leads to pessimism about deontological explanations. However, in this chapter, I shall treat these failures as opportunities to better understand what deontologism would have to look like, if it is to have any chance of explaining accessibilism.

The chapter is structured as follows. I relate Ginet's explanation in §3.2. In §3.3, with help from William Alston, we shall see that the explanation fails for two different reasons. But for each failure, we shall draw a lesson about what deontologism would have to look like, if it is to explain accessibilism. By the end of §3.3, we shall see that deontologism should be understood in terms of responses to what I called *potentially motivating reasons* in §2.5.1. Then, in §3.4, I continue in the theme of drawing out lessons about what deontologism would have to look like. We shall learn that deontologism should be understood in terms of the rationality of *forming* beliefs rather than the mere having of them, and that it should be cast in terms of a particular *kind* of blame, a kind of blame I will call *high-octane*. I conclude in §3.5.

§3.2   Ginet's Deontological Explanation of Accessibilism

Let us begin with Ginet's formulation of deontologism:

> One is justified in being confident that *p* if and only if it is not the case that one ought not to be confident that *p*: one could not be justly reproached for being confident that *p*.[66]

As we shall throughout the dissertation, let us treat 'justified' and 'rational' as interchangeable. The idea here is that believing *p* is rational for S just in case doing so is blameless for S. So, the above qualifies as a statement of

> *deontologism*   Believing *p* is rational for S if and only if believing *p* is blameless for S.

Ginet, I should clarify, is concerned with the property of its being rational to believe *p*, not the property of rationally believing *p*. It might be rational for S to believe *p* even if S does not believe *p*, and even if S believes *p*, but not rationally. This is commonly known as the distinction between *propositional* rationality, namely its being rational to believe, and *doxastic* rationality, which S has when S rationally believes.[67] That is, Ginet is concerned with propositional rationality. As well he should, since he is an accessibilist, and accessibilism about propositional rationality is more plausible than accessibilism about doxastic rationality. Notice that doxastic rationality requires propositional rationality plus, at least on a standard view, the right kind of causal relationship between the propositional rationality and S's belief.[68] Thus, the view that doxastic rationality is determined by what is accessible implies the bold claim that whether this causal relation holds is determined by what is accessible. By contrast, accessibilism about propositional rationality does not imply that bold claim. Throughout this dissertation, we shall be concerned with propositional rationality.

In any case, in light of Ginet's formulation of deontologism, we see that he holds there to be conceptual ties between 'ought', in his sense, and blameworthiness. In particular, it is blameworthy for one to believe *p* just in case one ought not to believe *p*, and so, accordingly, it is

---

[66]   Ginet 1975, 27.
[67]   See, e.g., Turri (2010) and Wedgwood (2013b) for discussions of the propositional/doxastic distinction.
[68]   See, e.g., Wedgwood (2006, 661).

blameless for one to believe *p* just in case it is not the case that one ought not to believe *p*. Ginet

holds that once these conceptual connections are combined with an 'ought implies can' principle,

it follows that rationality is determined by the very facts "directly recognizable to S at that

time."[69] Direct recognizability is a kind of accessibility, where a fact is directly recognizable just

in case one needs "only to reflect clear-headedly on the question of whether or not that fact

obtains in order to know that it does."[70] That is, Ginet holds that deontologism implies what I

called (A)$^{\text{very things}}$ in Chapter 2. However, (A)$^{\text{very things}}$ implies accessibilism as defined in Chapter

2, i.e.,

> *accessibilism*   Whether believing *p* is rational for S is determined by what is accessible to
> S.

In light of this, it is possible to draw from Ginet's explanation a (purported) explanation for

accessibilism as we understand it. This is what I shall do.

> Here is how, according to Ginet, deontologism explains accessibilism:

>> Assuming that S has the concept of justification for being confident that *p*, S
>> *ought* always to possess or lack confidence that *p* according to whether or not he
>> has such justification. At least he ought always to withhold confidence unless he
>> has justification. This is simply what is meant by having or lacking *justification*.
>> But if this is what S ought to do in any possible circumstance, then it is what S
>> *can* do in any possible circumstance. That is, assuming he has the relevant
>> concepts, S can always tell whether or not he has justification for being confident
>> that *p*. But this would not be so unless the difference between having such
>> justification and not having it were always directly recognizable to S.[71]

Ginet begins by saying that one ought to believe *p* whenever it is rational to do so and to refrain

from believing *p* whenever it is not rational to do so. Perhaps realizing that from the fact that it is

blameless for one to do something, it does not follow that one *ought* to do it, Ginet weakens the

---

[69]   Ginet 1975, 32.
[70]   Ibid., 33.
[71]   Ibid., 36.

premise as follows: one ought to refrain from believing *p* whenever it is not rational for one to believe *p*.

Employing this weaker premise, let us represent the line of thought as follows, where we shall take (9) and (15) to be shorthand statements of deontologism and accessibilism, respectively:[72]

(9)   Rationality is blamelessness.

(10)   If rationality is blamelessness, one ought to refrain from believing *p* whenever it is not rational for one to believe *p*.

(11)   If one ought to do something, one can do it.

(12)   If one can refrain from believing *p* whenever it is not rational for one to believe *p*, one can tell, for any *p*, whether it is rational for one to believe *p*.

(13)   If, for any *p*, one can tell whether it is rational for one to believe *p*, then, for any *p*, it is accessible to one whether it is rational for one to believe *p*.

(14)   If, for any *p*, it is accessible to one whether it is rational for one to believe *p*, then rationality is determined by the facts about which things are accessible to the subject.

(15)   Therefore, rationality is determined by the facts about which things are accessible to the subject.   [from (9), (10), (11), (12), (13), and (14)]

§3.3   Two Lessons from Two Shortcomings

In this section, with help from William Alston, we shall see that Ginet's explanation fails in two different ways. But for each failure, we shall draw a lesson about what deontologism would have to look like, if it is to explain accessibilism.

§3.3.1 The First Lesson

The first shortcoming concerns

(12)   If one can refrain from believing *p* whenever it is not rational for one to believe *p*, one can tell, for any *p*, whether it is rational for one to believe *p*.

---

[72]   This formulation of Ginet's argument is inspired by Alston (1989, 217); however, it does differ in that I have explicitly made deontologism a premise, and have used my own terminology. I've also started at '(9)', as we already used '(1)' through '(8)' in Chapter 2.

There appears to be no good reason to endorse (12) unless it follows from a more general

principle. What is this principle? First, let us read (12) through the eyes of a deontologist: If one

can refrain from believing *p* when it is *blameworthy* for one to believe *p*, one can tell, for any *p*,

whether or not it is blameworthy for one to believe *p*. So, the presupposed general principle is

the following:

> Can Tell   If one can avoid doing something when it is blameworthy for one to do it,
> then, for anything one could do, one can tell whether or not it is
> blameworthy for one to do it – that is, one can form the belief that it is
> blameworthy if it is, and the belief that it is not blameworthy if it is not.

The problem with (12) is that Can Tell appears to be false.

Consider the example of Huck Finn, from Mark Twain's classic.[73] Huck Finn has been

instructed from early on (we might prefer to call it brainwashing) that slaves are property, and so

that aiding a slave's escape is wrongful theft. Even so, Huck helps his good friend Jim, a slave,

escape. He is later approached by slave hunters. In Twain's story, Huck believes it is

blameworthy for him to hide Jim. But he does so anyways, lying to the slave hunters.

Of course, it is not blameworthy for Huck to hide Jim. In fact, it is plausible to me that it

would be blameworthy, at least to some degree, for Huck to *not* hide Jim. Indeed, Huck is aware

of conclusive reasons to hide Jim: that Jim is his friend, that Jim needs to see his wife and

children, and so on. Moreover, Huck can respond to these reasons, as he in fact does. In this case,

Huck can avoid the blameworthy alternative of not hiding Jim even though, presumably, he

cannot tell whether or not this alternative is blameworthy. He cannot form the true belief that it is

blameworthy to not hide Jim. In light of this, Can Tell is false.

In fact, the counterexample to Can Tell can be made more plausible. Suppose that rather

than straight-up believing that it is blameworthy for him to hide Jim, Huck merely finds himself

---

[73]   Twain 1885, Ch. 16. For philosophical discussions of the case of Huck Finn, see Bergmann (2006, 91) and Sripada (2015, 38-39).

– quite understandably – in suspension about the question, unable to form a true belief about whether it is blameworthy to hide Jim. At least, he is unable in the relevant sense of 'able'. Perhaps Huck is physically able to find and open a book on his raft to a certain page, where the contents of that page would so illumine that he would be able to form the true belief that hiding Jim is not blameworthy. But this is not the relevant sense of 'able'. It is the same sense in which I am "able" to open a safe whose combination I do not know, because I can physically move the dial in the relevant ways. At all events, in the relevant sense Huck cannot tell whether or not it is blameworthy to not hide Jim, being unable to form the true belief about this question. But even so, it would be blameworthy, at least to some degree, for him to not hide Jim, given the conclusive reasons to hide Jim of which he is aware. Can Tell is false. So, there is no good reason to endorse (12).

What can we learn from this shortcoming of Ginet's explanation? According to Can Tell, to be able to avoid a blameworthy alternative, one must have access to a reason which can guide one to avoid it. So far, so good. But, more specifically, Can Tell says we must have access to the following reason: *the fact that the alternative is blameworthy*. And this is a mistake. That it is blameworthy to not hide Jim is indeed a reason to hide Jim which could, in principle, guide Huck to hide Jim. But reasons which can guide Huck to avoid the blameworthy alternative need not be reasons *with that structure*. We should be much more liberal about reasons that can do this job. This is the lesson, then: to be more liberal about the reasons that can qualify as guides that enable one to avoid blameworthiness.

Plausibly, if one has the relevant ability to avoid doing something when it is blameworthy, then it is not by mere fluke or lucky happenstance that it is blameworthy.[74] This

---

[74]   I say this is plausible. And it is. But that does not keep it from being controversial. For a place to start, see Nagel 1993.

means that when one does something blameworthy, or blameless, one exercises a capacity that can be, in some sense, *guided* by certain things. We can give the name 'reasons' as placeholders for these things. We have learned that we should be more liberal than Can Tell suggests, with respect to what reasons, in this sense, can be.

We have also learned that deontologism should be understood in terms of the rationality of responses to epistemic *reasons*, in the sense of reasons just discussed:

deontologism It is rational for S to respond to epistemic reasons in a given way at *t* if and only if it is blameless for S to respond in that way at *t*.

§3.3.2 The Second Lesson

William Alston raises the second shortcoming with Ginet's explanation. Reflect on the fact that accessibility is a technical notion within accessibilism. For present purposes, the important point is that accessibility involves the *immediate* availability of the thing accessible to one. For example, that a piece of information can be found at one's local library by no means implies that it is accessible to one in the relevant sense.

So, then, examine the component of Ginet's explanation that first draws the connection to accessibility, namely

(13)   If, for any *p*, one can tell whether it is rational for one to believe *p*, then, for any *p*, it is accessible to one whether it is rational for one to believe *p*.

Alston raises the following problem. Even if being able to refrain from believing when it is not rational implies that one can tell whether it is rational for one to believe, there may be situations in which one can come to discern this precisely *because*, and *only* because, one can find information available at one's local library.

Alston gives a non-epistemic example to illustrate the point.[75] Consider a professor deliberating about whether or not it is blameless to take a job at another university for the fall without telling her current department until April 12th. She may not be able to tell whether it is blameless by mere clear-headed reflection; at the same time, she may be able to tell precisely because, and only because, she can research the university's policy on the matter at the library.

In general: contra (13), being able to tell whether something is so does not imply that whether it is so is accessible to one, in the sense relevant to accessibilism.

What can we learn from this shortcoming of Ginet's explanation?

Now (13) itself is framed in terms of the general principle I called Can Tell, according to which the ability to avoid blameworthiness requires that one can form a true belief about whether what one may do is blameworthy. The general idea behind (13) is that one can tell whether what one may do is blameworthy only if it is accessible to one whether what one may do is blameworthy. Having abandoned Can Tell in §3.3.1, however, our present interest should be in a slightly different general idea, namely the idea that a reason is a guide that enables one to avoid blameworthiness only if the reason is accessible to one. This general idea is the link we would need, to get from deontologism to accessibilism.

What is noteworthy is that Alston's problem for (13) raises a structurally similar problem with this link from deontologism to accessibilism. For suppose that it is on record at the professor's library that university policy has it that, before taking another job, one must give notice to one's current departure well before April 12th. Is not the fact that this is the university's policy, though, a reason that can guide the professor to give notice in advance, and so avoid appropriate blame? It seems so. Therefore, so the worry goes, it is false that a reason is a guide

---

[75]   Alston 1989, 217-218.

that enables one to avoid blameworthiness only if the reason is accessible to one, and the link to accessibilism is cut off.

But the link to accessibilism need not be cut off. Instead, we need to ask ourselves: in terms of *what kind* of obligation should we understand deontologism?

Suppose that the professor, before going to the library and learning the university's policy, has an obligation to tell the university well in advance of April 12th. Suppose also that she would not have this obligation in the first place if it were not university policy that the university be forewarned. What is the structure of this obligation?

In saying that she has this obligation, what are we saying it is appropriate to blame her for failing to do? Prior to learning about the policy, it would not be appropriate to blame her for failing to decide, *right then*, to forewarn her university. Our blame would be jumping the gun in that case. Instead, at this point we can only appropriately blame her for failing to go to the library or, perhaps, for failing to be concerned enough about the fact that she is unsure whether switching jobs without telling her university is on the up and up. It is only after the policy is a reason she can respond to that it would be appropriate to blame her for failing to forewarn her university. So, before she can respond to the policy itself, her obligation to forewarn the university has a *non-basic* structure: it is an obligation she has precisely because she has other obligations, which if fulfilled would link with others, which if fulfilled will result in her forewarning the university.

So, what we learn is that deontologism must be understood in terms of *basic* obligations – that is, obligations we have at the time, but not because we have any other obligations at the time. Equivalently, the reasons that are guides enabling one to avoid blameworthiness, to which deontologism refers, must be reasons that *can move the subject at the very time in question*. That

is, to recall §2.5.1, these reasons must be *potentially motivating reasons*. By definition, potentially motivating reasons *can* motivate one at the time, in an immediate sense of 'can'.

To restate our second lesson, epistemic reasons – i.e., the guides enabling one to avoid epistemic blameworthiness, to which deontologism refers – must be *potentially motivating* reasons, in particular. For all Alston's objection to (13) tells us, there might yet be a link between deontologism, so understood, and accessibilism. For it might yet be that something can potentially motivate one at the relevant time, in such a way that one could be blameworthy for how one responds to it at that time, only if it is accessible to one.

It is worth pointing out, though, that to understand deontologism in terms of responses to potentially motivating reasons is *not*, in itself, to understand deontologism in terms of responses to reasons *accessible* to one. The connection to accessibility still needs to be shown. Indeed, in principle, potentially motivating reasons could be nearly *anything*: the object of an unconscious mental state, the content of a mental state that is part of an available reliable belief-producing process, which may be a fact one is entirely unaware of, yet which makes that process available, and so on. Perhaps reliabilists and other externalists lack practice in talking about potentially motivating reasons as we understand them, but this is an artifact of how they label the moving parts of their theory, and not for any principled reason. In any case, construing deontologism in terms of responses to potentially motivating reasons does not beg any questions against reliabilists and other non-accessibilists. A connection to accessibilism still needs to be shown.

So, we have internalized two lessons from two shortcomings of Ginet's explanation. These lessons have taught us what deontologism would have to look like, if it is to explain accessibilism. In the next section, I continue in this vein, and further clarify what deontologism would have to look like.

§3.4   Further Lessons

The above two lessons give us a better understanding of deontologism. Deontologism is to be understood as follows,

> *deontologism*   It is rational for S to respond to epistemic reasons in a given way at *t* if and only if it is blameless for S to respond in that way at *t*.

where epistemic reasons are understood as potentially motivating reasons.

In light of this, we can unpack accessibilism, as clarified in Chapter 2, in the following way:

> *accessibilism*   Whether it is rational for S to respond to epistemic reasons in a given way at *t* is determined by the facts about which things are accessible to S at *t*.

With these definitions on board, we can now state the principle that needs to be true, if deontologism is to explain accessibilism. It is the following:

> The Bridge Principle   Whether it is blameless for S to respond to epistemic reasons in a given way at *t* is determined by the facts about which things are accessible to S at *t*.

The conjunction of deontologism and the Bridge Principle imply accessibilism.

In this chapter, we have been drawing out lessons about what deontologism would have to look like, if it is to have a chance of explaining accessibilism. Let us continue in that theme. There are several respects in which our definition of deontologism can be made more specific. And it turns out that, on some perfectly intelligible ways of making deontologism more specific, the Bridge Principle becomes immediately implausible. So, in what follows, I shall specify how we need to understand the view, in order to give the Bridge Principle a fighting chance.

In particular, I shall focus on two questions. First, are responses to epistemic reasons to be thought of as the mere *having* of belief states, or instead as the *forming* of them? Second, what *is* blamelessness, in the first place?

If we want the Bridge Principle to be plausible, we should hold that responses to epistemic reasons are formings of beliefs. For consider the plausibility of a Bridge Principle in terms of merely having belief states. On that Principle, the facts about whether it is blameless to *have* a belief state at *t* are determined by the facts about which things are accessible to one at *t*. Now, it is far from obvious that we have, in the first place, obligations that range over what states we are in at a time, as opposed to over the events or processes that take us into and out of these states.[76] But even if we think such obligations exist, it is quite implausible that the facts about whether it is blameless to have a belief state at *t* is determined by the facts about which things are accessible to one *at that time t*. For example, suppose Beth, in her early forties, has had the following two belief states since she was six: the belief that broccoli is good for one's health, and the belief that Anchorage is the capital of Alaska. (By the way, Juneau is the capital of Alaska, not Anchorage.) At *present*, there might be nothing accessible to Beth that is in any way relevant to whether it is appropriate to hold either of these beliefs: she has not had cause to think about these questions for decades, and is not thinking of them now. So how can the facts about which things are accessible to Beth *now*, *at this very moment*, determine whether she is fulfilling her obligation now, at this very moment, in holding either of these beliefs? There appears to be no easy answer to this question.

To see this more clearly, on the assumption that there are obligations ranging over states, it must make sense to say that, in a case like Beth's, Beth violates no obligation in having the belief that broccoli is good for one's health, but does violate an obligation, and so is blameworthy, in having the belief that Anchorage is the capital of Alaska. There seem to be two ways this could work. Perhaps Beth was blameworthy in forming the belief that Anchorage is the

---

76   See Podgorski (2016a and 2016b).

capital of Alaska, and the blameworthiness of her having this state is inherited. Or, perhaps, the belief that Anchorage is the capital of Alaska does not cohere with Beth's other beliefs. But on the first view, while it might be facts about which things were accessible to Beth *in the past* that determine whether Beth was blameless in *forming* the belief, it is *not* facts about which things are accessible to her *in the present* that determine whether she is blameless in *having* the belief. And similarly on the second view. Beth may be sound asleep, such that nothing is accessible to her at the time. Nevertheless, it might be that her belief about Anchorage fails to cohere with the other beliefs she has at the time. But clearly, it is not the facts about which things are accessible to her at the time that, on either view, determine whether she is fulfilling her obligation.

So, the Bridge Principle is plausible only if we understand responses to epistemic reasons to be formings of beliefs, not the mere having of them. Therefore, we shall henceforth understand responses to epistemic reasons in this way.

Let us turn to our second question: what *is* blamelessness, in the first place? Of course, in one sense of 'blameless', something is blameless even when no obligations apply to it in the first place, as when we say that we cannot blame the stars for twinkling as they do, or my stomach for digesting food as it does. But recall from §1.2.3 that 'blameless', in this dissertation, refers to the fulfilling of an obligation; hence, an obligation must apply to one, in the first place. And in some sense of 'can', an obligation applies to one only if there is both something blameworthy one *can* do at the time as well as something blameless one *can* do at the time instead. In light of understanding deontologism in terms of responses to epistemic reasons, we see that the relevant kind of blamelessness – and blameworthiness, for that matter – is always a matter of responding in one way to epistemic reasons rather than another way available to one at the time.

Connected to this last point, in the sense at issue in this dissertation, blamelessness or blameworthiness for a response at a time is always *original*. That is, the property always attaches

to what one does *at the very time in question*. Whether one has the property at the time cannot derive from the fact that one has the property at some earlier time.[77] To illustrate, suppose a driver swerves onto a sidewalk, killing a pedestrian. But suppose also that, being drunk, he was unable to have driven more carefully. The drunk driver is blameworthy for killing the pedestrian because, in a more fundamental sense, he is blameworthy for having gone to the bar knowing he would drive home drunk. He is derivatively blameworthy for killing the pedestrian, but originally blameworthy for going to the bar.

So far, we have addressed a few structural questions about what 'blamelessness' refers to. We know that it is the avoiding of blameworthiness. But what is *blame*, in the first place? And what is it to be worthy of it? We shall need to get clear on these questions, for some answers to them leave the Bridge Principle un-guardable.

Let us begin with the question of what *blame* is. Some have held that blame is merely a cognitive attitude, a kind of belief that simply grades someone along some normative dimension, e.g., that what they did was bad or that they have an irritable disposition.[78] But this view has its limits. Such grading beliefs are not sufficiently distinct from beliefs about another's physical beauty, raw intelligence, or athleticism. Blaming someone is entirely different from judging that they are un-athletic, for example. And unsurprisingly, many have thought that what makes for the difference is that blame is at least partly *affective* or *emotional*.[79] For example, we cannot feel appropriately angry with someone merely for being un-athletic, but it may be possible to feel appropriately angry with someone for having acted inconsiderately. I will go along with this common assumption that blame, at least in the interesting sense, is at least partly affective or emotional.

---

[77]   For the distinction between original and non-original blameworthiness, see, e.g., Fischer and Ravizza (1998, 49-50).
[78]   See, e.g., Smart 1961.
[79]   See, e.g., Strawson 1962.

Even with this clear, there are several different kinds of blame worth distinguishing. I will distinguish them in terms of what I shall call degrees of "octane". The degrees of octane correspond to the kind of abilities or capacities it is presupposed that the objects of blame have. The higher the octane, the more sophisticated or substantive are the abilities presupposed. I distinguish four kinds of blame:

No octane: blame as mere attribution of causal responsibility. When we say that the atmospheric low pressure is *to blame* for the temperature, we mean merely that the low pressure is causally contributory to the temperature and express our frustration at how undesirable the temperature is.

Low octane: shame and analogues. It is possible to feel *shame* at being, say, un-athletic, even with the assumption that the lack of athleticism is entirely genetic. I suspect there are analogues of shame that are other-directed. Though common enough, it is not clear to me that shame is ever appropriate, except perhaps in situations where it serves as a corrective to feeling overconfident with respect to some kind of raw capacity one may have, or take oneself to have.

Medium octane: disapproval. Disapproval is a negative emotion directed at one for doing something it is not reasonable to expect her to have avoided doing at the time, yet which is the kind of thing a person, and in particular her future self, could be expected to avoid doing.

To illustrate, consider the case of Bobby the toddler. Bobby is sitting in the sandbox, and another toddler accidentally kicks sand onto Bobby. Suppose that the appropriate reactions to being accidentally kicked with sand differ from the appropriate reactions to being intentionally kicked with sand – for example, that while in the latter case it is appropriate to express a degree of anger, in the former case it is not. But Bobby cannot yet distinguish accidental from intentional kickings of sand. Moreover, it is unclear whether he would have reacted differently even if he could, due to his yet tenuous control over his temper. In any case, upon being

accidentally kicked with sand, Bobby becomes quite angry and calls the other toddler a "meanie".

Now, in the sense of 'inappropriate' at issue, it may not be reasonable to think Bobby could have avoided doing something inappropriate at the time. He is not yet able to interpret others' intentions, or control his temper, each of which seem necessary for him to be in a position, at the time, to respond appropriately to such a situation. However, it is clear that responding appropriately to future situations like this is the kind of thing Bobby could, with maturation and guidance, learn to do.

Since there is a sense in which Bobby could not have responded any better at the time, there are certainly some affective responses to Bobby that are not appropriate. However, the case is a paradigm of a case in which *disapproval* is appropriate. Disapproval at calling the other toddler a "meanie" is the kind of emotion Bobby's parents should communicate to him, to give him a sense that he is expected to try to respond differently next time, to pique his interest in the perspective of the other child, etc. The emotion is meant to help Bobby see doing better next time as something that matters, as a problem he may need to play some role in helping solve. Disapproval can be self- as well as other-directed.

High-octane: guilt, resentment, and indignation. Guilt is first-personal, resentment second-personal, and indignation third-personal.[80] High-octane blame is appropriately directed at a person only if it *is* reasonable to expect them to have not done what they did. This is why it is not appropriate for the parents to feel indignation, rather than merely disapproval, at their toddler Bobby. Similarly, in the intrapersonal case: appropriate self-disapproval may obtain in situation where appropriate guilt may not. In what sense could the person have done otherwise, if high-

---

[80]   For discussion of guilt, resentment, and indignation, see Strawson 1962 and Wallace 1994.

octane blame is to be appropriate? This is controversial. The debate about the compatibility of blameworthiness and determinism is, as I hope is clear, best understood as a debate about the compatibility of high-octane blameworthiness and determinism. What incompatibilists think determinism threatens is high-octane blameworthiness. And compatibilists think determinism is compatible with high-octane blameworthiness, insofar as they are not changing the subject.

Before turning to the question of what kind of blame we should take to be at issue in deontologism – if we are to expect the view to explain accessibilism – let us say what it is for blame to be appropriate. Some hold that emotions represent the world as being a certain way.[81] On this view, fear of the approaching lion represents the lion as dangerous, mournful sadness represents the loss of something good, and etc. If we endorse this view about emotions, we can say that an emotion is appropriate just in case the world is the way the emotion represents it as being. Then, to say blame is appropriate is like saying a belief is true. I shall not commit to this picture of emotions and so, not commit to this picture of blameworthiness. But take the case of fear. Regardless of whether we think of fear as literally representational, I suspect we can all agree that there is a certain sense of the word 'fit' such that fear *fits* a situation if and only if the situation involves danger. Such situations are fear-worthy. And the point generalizes. So, blame is appropriate just in case it fits. Of course, to forestall confusion, to say blame fits the situation is not necessarily to say that we should blame the person *all things considered*. Just as there may be many reasons it would be inadvisable to fear even in the presence of danger, there may be many reasons it would be inadvisable to blame in a particular case, even if blame fits the situation.[82]

---

[81]   See, e.g., Johnston 2001 and Doring 2007.

[82]   For this point, see Rosen (2004, 297) and Fischer and Ravizza (1998, 7).

In any case, blameworthiness is the property something has when blame directed at it would be appropriate. And blame is appropriate just in case it fits, in the sense discussed above. In saying this, I do not take myself to be saying something very substantive. The word 'fits' is almost a synonym of 'appropriate'. I intend merely to have dusted off and clarified what 'blameworthiness' means. I have not given a substantive account of the property to which the word refers.

In any case, it is time to see why the viability of the Bridge Principle depends on what kind of blame, and so what kind of blameworthiness and blamelessness, is at issue. Consider the following version of the Bridge Principle:

the Bridge Principle^{medium octane}

Whether it is *medium-octane* blameless for S to respond to epistemic reasons in a given way at *t* is determined by the facts about which things are accessible to S at *t*.

From reflecting on the case of Bobby in a general way, we can see that the Bridge Principle^{medium octane} is implausible. In the example in which another toddler accidentally kicks sand on him, it is a fact that it is medium-octane blameworthy for Bobby to respond by calling the other toddler "meanie". But now consider a variation of the case in which the kicking was intentional, but yet in which *the facts about which things are accessible to Bobby are exactly the same*. It might well be, we can suppose, that in this second variation of the case, it is in fact medium-octane *blameless* for Bobby to respond by calling the other toddler "meanie". (The other toddler was mean, after all.)

It is plausible to suppose that the medium-octane blameworthiness-statuses of Bobby differ in these two variations of the case. When the question is whether it is appropriate to disapprove of Bobby, we do not treat Bobby's present perspective, i.e., the facts about which things are accessible to him, as fixing whether he is blameworthy. We are also on the lookout for the environment Bobby is in. And part of what we want to teach Bobby, in expressing

disapproval of his action, is about the cases in which his perspective isn't suited to the environment he is in, whether he is aware of this or not. Part of what we want to teach him is to have a different, expanded perspective next time.

Since the Bridge Principle^{medium octane} is implausible, the Bridge Principle is also implausible if we take the kind of blame at issue to be low-octane, or zero-octane. Our only hope, if we want the Bridge Principle to have any plausibility, is to understand it in terms of *high-octane* blame and blameworthiness. We need to understand it in terms of the appropriateness of our most serious reactive emotions: guilt, resentment, and indignation.

With that said, we have learned our last lesson about what deontologism must look like, if it is to explain accessibilism.

§3.5   Conclusion

In Chapter 2, we used the regress objection as an opportunity to clarify our understanding of accessibilism. Similarly, in this chapter, we have used the failures of Ginet's deontological explanation as opportunities to clarify deontologism. Now, with accessibilism and deontologism clarified, we know the Bridge Principle that needs to be true, if deontologism is to explain accessibilism. Therefore, with the necessary scaffolding in place, it is time to ask why the Bridge Principle might be true. In the next chapter, I offer a proposal on this front.

disapproval of his action, is about the cases in which his perspective isn't suited to the environment he is in, whether he is aware of this or not. Part of what we want to teach him is to have a different, expanded perspective next time.

Since the Bridge Principle$^{\text{medium octane}}$ is implausible, the Bridge Principle is also implausible if we take the kind of blame at issue to be low-octane, or zero-octane. Our only hope, if we want the Bridge Principle to have any plausibility, is to understand it in terms of *high-octane* blame and blameworthiness. We need to understand it in terms of the appropriateness of our most serious reactive emotions: guilt, resentment, and indignation.

With that said, we have learned our last lesson about what deontologism must look like, if it is to explain accessibilism.

§3.5   Conclusion

In Chapter 2, we used the regress objection as an opportunity to clarify our understanding of accessibilism. Similarly, in this chapter, we have used the failures of Ginet's deontological explanation as opportunities to clarify deontologism. Now, with accessibilism and deontologism clarified, we know the Bridge Principle that needs to be true, if deontologism is to explain accessibilism. Therefore, with the necessary scaffolding in place, it is time to ask why the Bridge Principle might be true. In the next chapter, I offer a proposal on this front.

Chapter 4: Why Blameworthiness Requires Consciousness

§4.1   Introduction

In Chapters 2 and 3, we clarified deontologism and accessibilism. This has clarified that

the principle which would bridge deontologism to accessibilism is the following:

> The Bridge Principle Whether it is blameless for S to respond to epistemic reasons in a given way at *t* is determined by the facts about which things are accessible to S at *t*.

In this chapter, I develop an explanation of the Bridge Principle.

I shall proceed by offering an explanation of the view that one is blameworthy only for

responses to reasons of which one is conscious. This kind of view has an intuitive source.

Consider the example of Kenneth Parks. Early in the morning on May 24th, 1987, Parks drove to

his mother-in-law's home and stabbed her. He later arrived at a police station with cuts on his

hands. In the course of his trial, it became clear he had been sleepwalking. He was acquitted of

murder, with the following legal defense:

> It is generally accepted that sleepwalking is a state of automatism in which an individual is unaware of, and has no control of, his or her behavior.[83]

The reasoning is that Parks is guilty of murder only if he is blameworthy, and that he is not

blameworthy because he was not conscious of any reason to stop what he was doing at the time

of the killing. This suggests that, in this part of criminal law, it is assumed that one could be

blameworthy only for one's responses to reasons of which one is conscious. This assumption,

moreover, dovetails with our intuition. Before learning Parks was sleepwalking when he killed

his mother-in-law, it is easy to blame him. Afterwards, it is not so easy.

---

[83]   Broughton and Billings, et al, 1994, 254.

Of course, without an explanation of *why* blameworthiness should require consciousness, it is unclear how seriously we should take our intuition that it does.[84] Compare to the intuition that blameworthiness requires indeterminism. Without an explanation, it is unclear how seriously we should take that intuition, either.[85]

The kind of explanation I shall developed is foreshadowed by Neil Levy. Levy, appealing to what cognitive science has to say about consciousness, develops two different explanations of why blameworthiness should require consciousness.[86] On the first, consciousness is required because otherwise one's action does not express one's evaluative attitudes. On the second, consciousness is required because otherwise one's action is not an exercise of control. The explanation I shall develop is of the second kind, i.e., it is a *control-based* explanation. Given the connection Parks' legal defense draws between one's being "unaware of" and one's having "no control of" one's behavior, control is a natural place to start. However, my control-based explanation shall differ significantly from Levy's.

Levy's control-based explanation appeals to John M. Fischer and Mark Ravizza's notion of *guidance control*.[87] Roughly, guidance control is the capacity to respond to reasons in a flexible way. Levy argues that only consciousness can procure flexible responsiveness to reasons. I shall call this *the guidance explanation*.

As we shall see, it is controversial in cognitive science whether only consciousness can procure flexible responsiveness to reasons. So, whether the guidance explanation is sound depends on how the cognitive science turns out. Here we may draw a comparison to the compatibilism/incompatibilism debate. Compatibilists are wont to urge that we should not have

---

84   King and Carruthers (2012) attempt to explain this intuition away.
85   Murray and Nahmias (2014) offer a debunking explanation of the intuition that blameworthiness requires indeterminism.
86   Levy 2014, Ch. 5-6.
87   Fischer and Ravizza 1998.

to wait on physics' verdict about determinism to know we are sometimes blameworthy.[88] Similarly, perhaps we should not have to wait on cognitive science to know blameworthiness requires consciousness. Tellingly, when the court acquitted Parks, they did not wait on cognitive science. These reflections invite the question of whether a control-based explanation could be developed which is less vulnerable to scientific falsification.

According to my control-based explanation, the person herself must exercise control, which requires that she responds to reasons she herself recognizes. But the person herself, as opposed to a part of her, recognizes a reason only if she is conscious of it. With its emphasis on *the person*'s control, I shall call this *the personal explanation*.

Intriguingly, after developing the guidance explanation, Levy mentions, in a single paragraph, an additional explanation with some similarity to the personal explanation.[89] But as we shall see, Levy's additional explanation depends on the same scientifically controversial theses upon which the guidance explanation depends, while the personal explanation does not. Given common practice in criminal law, it seems worthwhile to explore explanations which are, in this way, less vulnerable to scientific falsification.

The chapter is structured as follows. §4.2 contains preliminaries, including clarification of what we shall mean by *consciousness*. In §4.3, I present intuitive data which supports the view that one is blameworthy only for responses to reasons of which one is conscious. I discuss the guidance explanation in §4.4. In §4.5, I present the personal explanation, and I respond to objections to it in §4.6. I extend this explanation to the Bridge Principle in §4.7. I conclude in §4.8.

---

[88]   See Strawson 1962.
[89]   Levy 2014, 114.

§4.2   Preliminaries

I begin by clarifying what is meant by *consciousness*. We say that one who wakes up has become conscious, attributing consciousness to the person *tout court*. But we also say, for example, that one is conscious of her surroundings, or of an impending deadline. Here, we attribute to the person consciousness *of* something in particular. I shall assume that consciousness *tout court* is the property of being conscious of something or other.[90]

As some use the word 'consciousness', to be conscious of something is for there to be something it is like for one to have the state which has this thing as its object. To give examples: there is something it is like for me to have a visual experience of my computer, or an auditory experience of Beethoven's *Moonlight Sonata*. And I should mention that, consonant with a recent trend in the philosophy of mind, I have no problem with the view that there is something it is like for me to entertain the next proposition I shall attempt to express.[91] There can also be something it is like to be aware of something via memorial experience, occurrent factual recall, occurrent belief, and so forth. Thomas Nagel dubs consciousness, so conceived, *phenomenal consciousness*.[92]

Levy does not (first and foremost)[93] have phenomenal consciousness in mind, when he speaks of consciousness. Instead, Levy refers to an occurrent state whose propositional content "the agent is able to effortlessly and easily retrieve…for use in reasoning".[94] An occurrent state is a state playing its distinctive role. For example, when I am answering the question 'What is the capital of California?', my belief that Sacramento is the capital of California is occurrent. But

---

[90]   Willard (1967, 513-523) and Hopp (2011, 13-18) have argued we can be conscious of things that do not, in fact, exist. I am neutral on this question.

[91]   See Pitt 2009, Horgan and Tienson 2002, and Bayne and Montague's (2011) anthology on cognitive phenomenology.

[92]   Nagel 1974.

[93]   Levy (2014, 28-29) is neutral with respect to accounts on which 'phenomenal consciousness' and 'consciousness' as he shall understand it, though conceptually distinct, actually refer to the same metaphysical kind.

[94]   Levy 2014, 33.

normally, this belief state is merely dispositional, requiring a special trigger or context for it to play its distinctive role.[95] The content of an occurrent state is easily retrievable for use in reasoning when no special promptings are needed for one to put it to use in reasoning. To have an occurrent state whose content meets this condition is, for Levy, to be *conscious* of that content. Essentially, consciousness so conceived is what Ned Block calls *access consciousness*.[96]

In this dissertation, I shall be concerned with access consciousness. Primarily, this is because my explanation for why blameworthiness requires consciousness has most plausibility, and appeals to the widest audience, when cast in terms of access consciousness.

However, it shall be helpful to keep in mind the following. At times in this dissertation, I have found it dialectically helpful – though not, I think, essential to my central argument – to speak of what things are like for the subject, phenomenologically, and then transition seamlessly to talk about her consciousness. When I do this, I still refer to access consciousness. It is just that I am relying on the assumption that if one is phenomenally conscious of something, then she is access conscious of it, too. Indeed, how could one be phenomenally conscious of something, yet fail to be poised, in principle, to use that object in reasoning and action? Being poised in this way does not require the object's being in focus, or central to what one is up to at the time. In any case, it is fair to say that as a matter of empirical fact: generally, phenomenal consciousness implies access consciousness. So, for any particular occasion on which I rely on the assumption that this kind of connection holds, the reader is invited to try to show that that particular connection is out of step with the empirical facts.

---

[95] For the occurrent/dispositional distinction, see Audi 1994.

[96] With one caveat: For Block (1995, 231), one is access conscious of a content only if it is widely available to one's cognitive systems. And while, on Levy's definition, one is conscious of a content only if it is easily retrievable for use in reasoning, according to Levy it is an empirical matter whether contents retrievable in this way are widely available to one's cognitive systems. However, Levy does think this empirical question is answered in the affirmative, which makes Levy's notion of consciousness and access consciousness co-referential. See Levy 2014, 35-36.

And notice that to assume that phenomenal consciousness implies access consciousness is *not* to take a stand on whether philosophical zombies, i.e., access conscious beings who lack phenomenal consciousness, are possible.[97] Whether philosophical zombies are possible turns on whether access consciousness implies phenomenal consciousness. That is the converse of the assumption I have highlighted. And I shall take no stand on it.

At least, I shall take no *official* stand on it. In this paragraph, let me reveal my own, perhaps idiosyncratic, hand. I suspect that phenomenal consciousness metaphysically entails access consciousness but not vice versa. So, I suspect philosophical zombies are possible. Moreover, I suspect blameworthiness requires phenomenal consciousness, in addition to the access consciousness which, on my view, is entailed by phenomenal consciousness. But establishing these suspicions about the importance of phenomenal consciousness is outside the scope of this dissertation. Even on my assumptions, if phenomenal consciousness is required for blameworthiness, so is access consciousness. So if my idiosyncratic hand is correct, the view I argue for in the dissertation is logically weaker than the view that phenomenal consciousness is required for blameworthiness. No wonder, then, if it is more plausible and appeals to a wider audience.

With consciousness clarified, I should like to briefly remind us how I understand blamelessness in this dissertation. Blamelessness is the avoiding of presently available blameworthiness. And vice versa: blameworthiness is the avoiding of presently available blamelessness. So, one is blameless for doing something at *t* only if, at *t*, one can do something which is blameworthy to do and one can do something which is blameless to do. Blameworthiness, recall, is *original*: it does not derive in any way from past blameworthy

---

[97]   See Chalmers 1996.

doings. In light of this, both blamelessness and blameworthiness attach exclusively to responses to reasons capable of moving one at that very time, i.e., potentially motivating reasons.

§4.3   Intuitive Data

In this section, I briefly present intuitive data which supports the view that one is blameworthy only for responses to reasons of which one is conscious. None of this data is conclusive. But it is a nice appetizer: it helps us see why an explanation of this view would be nice to have.

Remember from §1.2.2 the *new evil demon problem* for reliabilism. Andrew Moon argues that a similar problem faces any view on which rationality is not fixed by the facts about which conscious states one has.[98] Thus Moon: Consider Augustine, author of *Confessions*. At any time in Augustine's life, he had many unconscious beliefs and other mental states, in addition to conscious states. On the other hand, consider Augustine*. At any time in Augustine*'s life, he only has conscious mental states. More specifically, a malicious demon has destroyed all the unconscious states Augustine* would have otherwise had, and at any time at which he appears to himself to retrieve an unconscious state, in fact the demon is creating a new conscious state from scratch. At any time, Augustine cannot, from the inside, distinguish his position from that of Augustine*'s position. According to Moon, it is intuitive that it is justified or rational for Augustine to respond to reasons in a certain way just in case the same holds for Augustine*.[99]

If we read this intuitive result through the eyes of deontologism, the upshot is that conscious states fix whether it is blameworthy to do something. But this strongly suggests that it is only responses to reasons of which one is *conscious* for which one could be blameworthy, in

---

[98]   See Moon 2012 and 2015.
[99]   Moon 2012, 349.

the first place. Otherwise, it is unclear why conscious states should fix blameworthiness in this way.

Now, Kevin McCain has objected to the idea that Moon's case is metaphysically possible, and Moon has replied by restricting the activity of the demon in various ways, yet with an eye to the same intuitive upshot.[100] While I do not think McCain's objections are decisive, they do highlight the fanciful nature of Moon's thought experiment. For this reason, it may be worthwhile to see whether the same kind of intuition can be generated without invoking demons.

It can. Consider the practice we employ when checking to see whether guilt is appropriate.[101] Our practice is to exercise *episodic memory*,[102] which presents, in a certain manner, the very things one was conscious of, as it is a kind of reliving of one's past consciousness of them. We then take note of the facts about which reasons we were conscious of at the time. Assuming our memory confirms we did the deed in the first place, we inculpate or instead exculpate ourselves depending on the facts about which reasons we were conscious of at the time, concluding guilt is appropriate just in case we were conscious of conclusive reason to not do what we did. Moreover, even if we sometimes also use other practices, like consulting wiser or more informed persons, the episodic memory practice enjoys a kind of priority. The other practices, at least generally, would neither free us from the conviction of guilt if we remember doing the deed with consciousness of conclusive reasons not to, nor convince us wholeheartedly of guilt if we remember either not doing the deed, or doing it without consciousness of conclusive reason not to.

---

[100]   McCain 2015, 99-100; Moon 2015, 109-110. It is blameless/blameworthy to A only if it is possible to A at the time. Now consider the case where to A is to form a belief. One issue is whether it is possible for one to form a belief without having, at the time, an interlocking web of unconscious background beliefs. McCain thinks it is not. Moon urges that even if this is the case, we can imagine that the demon destroys only certain unconscious states, leaving the needed background beliefs intact.

[101]   Guilt can be *caused* in many ways: the realization that what one did harmed another, the excoriations of another, one's own sense of falling short during and immediately after the relevant event, or even mental illness. The current issue is how we check to see whether guilt is appropriate, however it may be caused.

[102]   To be contrasted with *factual memory*. It is by factual memory that I remember that George Washington was America's first president, for example. It might even be by factual memory that I remember, by rote instead of by needing to work it out, that 8 times 8 equals 64.

Since the practice described bases conclusions about guilt-worthiness solely on the facts about which reasons one was conscious of at the relevant time in the past, it is reliable only if it is only responses to reasons of which one is conscious at the relevant time for which one could be blameworthy. If one could be blameworthy for responses to reasons of which one is not conscious at the time, the practice would not be reliable. Of course, one option is to admit that the practice is not reliable. But to the degree that admitting this is unattractive, we have another source of intuitive support for the view at issue.

Perhaps the strongest source of intuitive support is discernible when we pay attention to what it is like to engage in first personal deliberation about whether to A, where 'A' is a variable for something one can do.[103]

Imagine that you are one of two subjects deliberating about whether to A. You discover the other subject and, after discussing the matter of A-ing, each of you see that the facts about which reasons each of you are conscious of are precisely the same. The other subject remarks that, whatever else is true, if one of you could not be blamed if she A-ed in response to these reasons, then neither could the other. What the other has said is intuitive: it feels like common sense, and not a remark only someone in the grip of a theory would make. But then if this kind of remark is true in every case like the one described, it is only responses to reasons of which one is conscious for which one could be blameworthy.

Let us dig into this case a little. Suppose a third person, call her Abby, claims in your hearing that one of you has an unconscious state the other lacks and that, by virtue thereof, one of you has a reason against A-ing the other lacks. Suppose Abby, accordingly, claims that it is blameworthy for one of you to A, but not the other. Abby says nothing to indicate which of you

---

[103]   Here I build on an argument George Sher (2009, 59) puts in the mouth of his opponent.

is which. Could what Abby says be true? Doesn't it make it seem like a matter of *luck* whether it is blameworthy for you to A? But how could that be a matter of luck?

Recovering from some initial bafflement, you and your compatriot reason that Abby must have meant that if each of you deliberate assiduously enough before A-ing, one but not the other will become conscious of a reason against A-ing, because one but not the other will bring the relevant unconscious state to consciousness. This, of course, is consonant with the view that it is only responses to reasons of which one is conscious for which one could be blameworthy, as in this eventuality the facts about which reasons one is conscious of differ. Abby insists, though, that this was not her intent, and that this way of interpreting her claim is a mere artifact of failing to stipulate, in the case, the relevant time *t* at which it is blameworthy for one of you to A. She insists that at any time *t* in the interval remaining for you and your compatriot to A or abstain, one but not the other will have a reason of which she is not conscious, by virtue of which it is blameworthy for her to A at *t*. At this, you and your compatriot blankly stare and can only mutter that this does not seem fair.

The above case dramatizes how intuitive it is to think that it is only responses to reasons of which one is conscious for which one could be blameworthy. In any case, from Moon's new evil demon problem, to reflection on how we use episodic memory to check whether we are genuinely guilty, to reflection on first personal deliberation, we see that there are a variety of sources of intuitive support for this view. What could explain it?

§4.4   The Guidance Explanation

The conclusion of the guidance explanation is that one is blameworthy only if "conscious of the moral significance of their actions".[104] This is not exactly the view that it is only responses

---

[104]   Levy 2014, 37.

to reasons of which one is conscious for which one could be blameworthy. But it is a step in its direction. To be conscious of the moral significance of the action is to be conscious of something *bad* about the action.[105] For example, if Abby is conscious that the bike belongs to someone else, she is conscious of something bad about riding it away. She is thus conscious of the moral significance of doing so.

The guidance explanation begins with the premise that guidance control is required for blameworthiness. As Fischer and Ravizza define the notion, to exercise guidance control is to exercise a capacity which meets two conditions.[106] First, were one to exercise this capacity, one would recognize an "understandable pattern" of reasons that concern the proposed action.[107] Second, in some cases one would react to sufficient reason to do otherwise by doing otherwise.[108]

To illustrate, suppose Abby steals a phone from a stranger. Abby may not have recognized the harm this theft causes as a sufficient reason to do otherwise, and she certainly did not react to it. But suppose that, exercising the capacities she actually exercised to steal, she would in different circumstances recognize as reasons the fact that the victim of the theft is a family member, or in need of a phone due to a medical problem, and so on. That is, imagine there is an understandable pattern of reasons she would recognize. And suppose she would react to some of these reasons by not stealing. Then Abby's theft is an exercise of guidance control. By contrast, suppose Beth is just like Abby except that the only reason she would recognize as a reason to not steal is that the person with the phone has a yellow shirt on. The pattern of reasons Beth would recognize does not make sense from, as Fischer and Ravizza put it, "the wide

---

[105]   Levy 2014, 36-37.
[106]   Fischer and Ravizza 1998.
[107]   Ibid., 72.
[108]   Guidance control is entirely compatibilist-friendly: though it is essential to the relevant capacities that they are exercised in different ways in different relevantly possible worlds, these worlds need not share the same history. See Ibid., 38-39.

reflective equilibrium of the community".[109] So, unlike Abby, Beth does not exercise guidance control.

Levy's strategy is to show that guidance control requires consciousness of the moral significance of the action. Consider again the case of Parks. Parks was responsive to *some* reasons. For example, he was responsive to visual cues when he drove to his in-laws' home. But "action scripts" governed his responses.[110] Actions scripts are neurological algorithms for automatized and inflexible behaviors, such as pulling the clutch when switching gears. While governed by action scripts, Parks would not recognize an understandable pattern of reasons. For action scripts are too inflexible.

According to Levy, the needed flexibility requires *integration*, that is, the capacity to assess salient features of one's situation "for consistency and for conflict with" a broad range of one's attitudes.[111] For example, the thing in front of Parks is his innocent mother-in-law, a fact which makes stabbing it dreadfully bad to do. Parks must be able to assess this fact with, say, his opposition to harming people, not to speak of his love for his mother-in-law. Otherwise, Parks' responsiveness is inflexible, and he is unable to recognize an understandable pattern of reasons.

The final step of the guidance explanation is that integration requires consciousness. Neuroimaging studies indicate that consciousness correlates with the firing of neurons spread throughout a "global neuronal workspace".[112] And when these neurons fire, the "functional correlation and coherence in activity" of otherwise disparate and encapsulated regions of the brain increases sharply.[113] Other studies contrast subjects who complete one task while distracted by another "effortful mental task, such as keeping a number in mind" with control subjects who

---

[109]   Ibid., 72, footnote 15.
[110]   Levy 2014, 75.
[111]   Ibid., 113.
[112]   Ibid., 49.
[113]   Ibid., 51.

are undistracted.[114] The distracted subjects outsource parts of the original task to unconscious

processes, which results in less coherent approaches to the task as compared to the undistracted

subjects. All these results dovetail nicely with the "integration consensus" in cognitive science,

on which the unique functional role of consciousness is integration.[115] That is, consciousness

alone allows for integration.

To put the guidance explanation succinctly: Blameworthiness requires guidance control.

Guidance control involves flexible responsiveness to reasons. Flexible responsiveness to reasons

requires the integration of the moral significance of the action with a broad range of attitudes.

And this requires consciousness of the moral significance of the action.

This explanation has its critics. Chandra Sripada argues that flexible responsiveness to

reasons does not require the integration of the moral significance of the action with a broad range

of attitudes. According to Sripada, the mind has many *motivational modules*, each responsive to

a special domain of reasons. For example,

> …there is a 'fear' system that stores information about important threats,
> performs rapid inferences about when a situation constitutes a threat, maintains a
> store of action plans for what to do about frequently encountered threats, and so
> on.[116]

Other modules concern "food and nourishment", "norms and obligations", and so on.[117] The

output of a module is a simple signal with a degree of strength. Each signal has a valence

according to whether, relative to the reasons in the module's domain, a given option is to be

pursued or avoided. A "downstream aggregation mechanism" then adds up the strength of these

signals and selects an action.[118]

---

114   Ibid., 52.
115   Ibid., 40.
116   Sripada 2015, 40-41.
117   Ibid., 41.
118   Ibid., 42.

On Sripada's picture, there need never be a point at which the moral significance of the action is integrated with a broad range of one's attitudes. The modules individually contact reasons, but one may be unable to assess these reasons for coherence with one's attitudes. And while the modules generate signals which have a minimal kind of content, the aggregation mechanism to which these signals are sent is unable to assess their content for coherence with a broad range of one's attitudes. All the aggregation mechanism can do is add up the strength of the signals, thereby selecting an action. But even so, Sripada urges, actions generated in this way are flexibly responsive to reasons. When only one module operates, the action may well be inflexible. But often many modules, each "separately sensitive to a diversity of considerations", generate the action together.[119] And this suffices for flexible responsiveness to reasons. Integration of the moral significance of the action with a broad range of one's attitudes is not required.

Levy offers two responses to Sripada.[120] First, modules are evolutionary. So, there are no modules responsive to the many reasons which do not bear on survival fitness. But flexible responsiveness requires contact with such reasons. Second, modules are "likely" to make use of "proxies" for reasons which are "only rough and ready trackers" of them.[121] And this means they cannot yield flexible responsiveness.

I worry about Levy's replies. Plausibly, in addition to evolution, enculturation and habituation also generate modules. So we may have all the modules we need, after all. I also wonder why it should be likely that modules respond to proxies for reasons, not reasons themselves. In addition, without more information about the difference between reasons and proxies, it is unclear whether even consciousness itself responds to reasons rather than proxies.

---

[119]   Ibid.
[120]   Levy 2015, 68-69.
[121]   Ibid., 69.

And finally, it is unclear why responsiveness to only proxies should make for inflexibility. Facial expressions, for example, are only rough and ready proxies for others' emotional states. Even so, they facilitate flexible responsiveness to these states.

At all events, in cognitive science, it is an open question whether flexible responsiveness to reasons requires the integration of the moral significance of the action with a broad range of attitudes. And this is not the only juncture at which the guidance explanation must wait on cognitive science. For one, it might be that unconscious processes can facilitate integration.[122]

But perhaps we should not have to wait on cognitive science for an explanation of why blameworthiness requires consciousness. When the court acquitted Parks, they did not wait. So, in the next section, I shall develop an explanation which is less vulnerable to scientific falsification.

§4.5   The Personal Explanation

The guidance explanation begins with guidance control as a constraint on blameworthiness. Then it links guidance control to integration, and integration to consciousness. We have seen how these links are vulnerable to scientific falsification. In this section, I develop a different explanation. It also begins with control. But it appeals to a notion of control more general than guidance control, namely the capacity to respond to reasons one recognizes. The link to consciousness is then forged on the basis of the premise that only persons are appropriate targets of blame. Given this premise, the person herself must exercise control, and so the person herself must recognize and respond to reasons. And as I shall argue, the person herself, as opposed to a part of her, recognizes a reason only if she is conscious of it. With its emphasis on what it takes for *the person* to exercise control, I call this *the personal explanation*. I develop it

---

[122]   See Mudrik et al (2011) for an argument to this effect, and Levy (2014, 55-58) for a reply.

using what Levy would call "fairly traditional philosophical arguments".[123] I show that it is not vulnerable to scientific falsification in the respects in which the guidance explanation is vulnerable. Then in §4.6, I address objections to it.

While the conclusion of the guidance explanation is that one is blameworthy for an action only if conscious of its moral significance, the conclusion of the personal explanation is that *one is blameworthy only for responses to reasons of which one is conscious*. We shall understand reasons as *potentially motivating reasons*, as explained in §2.5.1. A *response* to reasons is something done for or, as the case may be, in spite of reasons. A response is an event, such as raising one's right hand, or voting for Carol.

Now, events can be individuated in different ways. Individuated coarsely, there may be a case in which raising one's right hand is identical with voting for Carol. But individuated finely, raising one's right hand is distinct from voting for Carol, even if doing so makes it the case that one has voted for Carol.[124] The personal explanation is consistent with either approach to individuating events. And as we shall see more clearly in 6.5, my larger goals in the dissertation are consistent with either of them, as well. On the coarse-grained approach, its conclusion is that one is blameworthy only for *the features* of responses which are based on, or present in spite of, reasons of which one is conscious.[125] Suppose raising one's right hand is identical with voting for Carol. Even so, this event may be a vote for Carol for reasons of which one is conscious, but a raising of one's *right* hand, rather than one's left, for reasons of which one is not conscious. In that case, one could be blameworthy for the event's being a vote for Carol, but not for its being the raising of one's right hand. On the fine-grained approach, the conclusion of the personal explanation is that one is blameworthy only for *precisely the events* which are responses to

---

[123]   Levy 2014, 135.
[124]   See Davidson (1963) for the coarse-grained approach, and Kim (1966) for the fine-grained.
[125]   Compare Sher (2009, 4-5).

reasons of which one is conscious. On this approach, raising one's right hand and voting for

Carol are distinct events, and one might be blameworthy for the latter and not the former.

Now let us dive into the personal explanation. It begins with the old idea that one is

blameworthy only for exercises of *control*. Fischer and Ravizza, who developed the notion of

guidance control, express nicely how plausible this idea is:

> It seems to be a basic presupposition embedded in the way we think about these matters that an agent must in some sense control his behavior in order to be morally responsible.[126]

The explanation begins with a notion of control which is simpler and more general than

guidance control. To illustrate this notion of control, consider an intuitive contrast between a

rock and an automatic door. The rock is a slave to gravity and anything which may push it

around. Intuitively, the rock exercises *no* control over what it does. But things are different for

the automatic door. The automatic door has a photoelectric cell which detects incoming

pedestrians, in response to whom the door opens. And when opening, the door is not merely

being pushed around in the sense in which the rock is merely being pushed around. Intuitively,

there is a sense in which the door is exercising control.

We should expect this kind of contrast between the door and the rock. Guidance control

involves flexible responsiveness to reasons. But a necessary condition on flexible responsiveness

to reasons is responsiveness to reasons, full stop. And as Fischer and Ravizza think of the notion,

responsiveness to reasons is a capacity to respond to reasons one *recognizes*, in some sense.[127]

Fischer and Ravizza do not directly say what recognition is supposed to be. But,

presumably, to recognize a reason is to detect or register it in some way. And for all Fischer and

Ravizza say, recognition can take many forms. It need not be conscious. Levy for example,

---

[126]   Fischer and Ravizza 1998, 13-14.
[127]   Ibid., 69-73.

speaks about how Parks recognized "the traffic light as a reason" to stop, even while sleepwalking.[128] Recognition need not be a mental phenomenon, either. In fact, the automatic door appears to recognize the pedestrian as a reason to open. As John Searle has put it, "We say, "The door knows when to open because of its photoelectric cell…"".[129] According to Alvin Goldman, we talk this way because the door "has a reliable mechanism for discriminating between something being before it and nothing being there".[130] Whatever the right account of recognition may be, the point is that the door can respond to reasons it, in some sense, *recognizes*. The rock cannot. So, the door satisfies a necessary condition on guidance control which the rock does not, namely having a capacity to respond to reasons it recognizes. And perhaps this difference between the door and the rock explains the contrast between them along the dimension of control.

This suggestion has much going for it. Intuitively, the door exercises control when it opens for pedestrians. That is, it exercises control when it recognizes the pedestrian via its photoelectric cell and responds by opening. The door does not exercise control when strong winds force it open. And the difference is that the door does not recognize local atmospheric states as reasons to open. Even if these states are, in some sense, reasons to open, none of the door's special mechanisms recognize them. This is why the door's response to the wind is no more an exercise of control than the rock's response to the wind. So, we have pinpointed what explains the intuitive contrast between the door and the rock along the dimension of control. The door has a capacity to respond to reasons it recognizes. The rock does not.

Admittedly, the control the automatic door has is quite minimal. For one, guidance control involves responsiveness to a broad pattern of reasons. The door lacks this kind of

---

[128]   Levy 2014, 112.
[129]   Searle 2004, 206.
[130]   Goldman 1976, 791.

responsiveness. For example, the door recognizes no difference between an ordinary pedestrian and one with an assault rifle. It will open to both the same, despite the fact that a pedestrian's wielding an assault rifle may well be a sufficient reason to not open. Even so, with its capacity to respond to reasons it recognizes, the door has *something* by way of control which the rock lacks.

My starting point is the intuitive claim that, to be blameworthy, one must exercise *at least* the kind of control the automatic door has and the rock lacks. That is, one must exercise a capacity to respond to reasons one recognizes. In fact, we can get more specific. Recall the distinction between *occurrent* and *dispositional* states, introduced when defining consciousness in §4.2. This distinction cuts across states of all kinds, including states of automatic doors. And notice that when the automatic door exercises control in opening for a pedestrian, its recognition of the pedestrian is not dispositional but instead *occurrent*. This recognition is playing its distinctive role, at that time, in the door's internal mechanisms. Thus, to exercise at least the kind of control the automatic door has and the rock lacks, one must exercise a capacity to respond to reasons one *occurrently* recognizes. This point shall prove important later. Hereafter, for brevity, I shall use 'recognition' to refer to occurrent recognition.

There is a gap between the view that one is blameworthy only for responses to reasons one recognizes and the view that one is blameworthy only for responses to reasons of which one is conscious. For recognition need not be conscious. But this view is an important first step. The next thing to see: while to recognize reasons is not necessarily to be conscious of them, to be conscious of reasons is, indeed, to recognize them. If the automatic door recognizes a reason to open when its photoelectric cell detects an incoming pedestrian, then, *a fortiori*, I recognize a reason to prepare my umbrella when I consciously see the incoming clouds. One way to put it is that recognition of reasons is a genus of which consciousness of reasons is a species, just as *animal* is a genus of which *horse* is a species. So, the question is whether there is any reason to

think blameworthiness requires not simply recognition of reasons in general, but that species of recognition of reasons which is consciousness of reasons.

Intriguingly, many different parts[131] of human persons exercise a capacity to respond to reasons they recognize. White blood cells recognize viruses as reasons to attack, and respond by doing so. The stomach recognizes entering food as a reason to excrete gastric juices, and responds by doing so. And on and on. Of course, it is never appropriate to blame these parts of the human person, just as it is never appropriate to blame an automatic door. This is true even if their responses to reasons constitute malfunctions. Even if the responses of white blood cells are causally responsible for serious health problems, neither resentment nor indignation directed at these cells is ever appropriate.

These reflections suggest that a part of a person is never blameworthy for what it does. They also suggest, as a corollary, that it is never appropriate to blame a person for something one of her parts has done, but not she herself.[132]

This should not be surprising. It is independently plausible that only *persons* are appropriate targets of blame. Suppose I find a scratch on my car. Thinking someone keyed my car, I may feel resentment. But suppose I later learn it was a deer which had scratched my car with its antlers. Given this information, it may be appropriate for me to feel annoyed, but not resentful. For what scratched my car was not a person, and so not an appropriate target of resentment. Or suppose I see someone appearing to give me the middle finger. I may feel resentment. But suppose later I learn that this person suffers from *alien hand* syndrome. This is a disorder where the sufferer is not "able (directly) to control the behavior" of one of her hands, which responds in sometimes unwanted ways to "perceptual information" and "environmental

---

[131]   'Part' shall always refer to a *proper* part.
[132]   Unless we have in mind indirect blameworthiness, as a heavy drinker may feel appropriate guilt for his kidney's malfunctions, because he could have stopped drinking.

contingencies".[133] Given this information, it may be appropriate for me to feel annoyed, but not resentful. For it was the person's alien hand which gave me the middle finger, not she herself. And since the person's alien hand is not itself a person, it is not an appropriate target of blame. Only persons are appropriate targets of blame.

Many have agreed that only persons are appropriate targets of blame. The view is often traced to John Locke, who said that 'person' "…is a forensic term appropriating actions and their merit".[134] For contemporary examples, observe Jim Stone's statement that "[p]ersons are *conceived* as responsibility bearers",[135] or Lynne Baker's statement that "[o]nly persons have moral responsibilities".[136]

Since only persons are appropriate targets of blame, it must be the person *herself*, and nothing less, which exercises the kind of control the automatic door has and the rock lacks. That is, the person herself must recognize and respond to reasons. Here we may have found a reason to think blameworthiness requires not simply recognition of reasons in general, but that species of recognition of reasons which is consciousness of reasons. For it might be that for the person herself to recognize a reason, rather than for one of her parts to do so, is for her to be *conscious* of it.

The case of Parks suggests as much. Recall that, while sleepwalking, Parks' behavior was governed by action scripts, i.e., neurological algorithms responsible for automatized and inflexible behaviors in response to stimuli. Plausibly, the part of his brain which implements actions scripts recognized reasons, in response to which his fateful behavior issued. But, intuitively, it does not seem that Parks *himself* recognized reasons, at that time. In fact, Parks had

---

[133]   Levy 2014, 84.
[134]   Locke 2009 [1690], 376.
[135]   Stone 1988, 530.
[136]   Baker 2007, 29.

a close, caring relationship with his in-laws.[137] If he himself had recognized the situation, he would not have stabbed his mother-in-law. For he would have immediately seen how contrary such behavior is to his own goals and self-image. But if the fairly sophisticated unconscious recognition obtaining in this case does not qualify as recognition by Parks himself, it begins to seem doubtful whether any kind of unconscious recognition could. This suggests that for a person herself to recognize a reason, rather than for one of her parts to do so, is for her to be conscious of it.

And the argument can be made tighter. Plausibly, essential to and distinctive of being a person is having a certain array of *rational capacities*. Without pretending to give a definitive list, these capacities include theoretical, moral, and practical reasoning. These kinds of reasoning in turn result in beliefs, intentions, and actions responsive to reasons. Presumably, the explanation for why non-human animals fail to qualify as persons is that they lack the relevant array of rational capacities.

Just as many have agreed that only persons are appropriate targets of blame, many have agreed that rational capacities are essential to and distinctive of persons. According to Boethius, "a person is an individual substance of a rational nature".[138] John Locke defined a person as "a thinking intelligent being that has *reason and reflection,* and can consider itself as itself…".[139] Immanuel Kant held that "beings without reason…are called *things*; rational beings, by contrast, are called *persons*…".[140] And for a contemporary example, consider Robin Dillon's statement that "to be a person is to be a rational being".[141]

---

137   Broughton and Billings et al 1994, 254.
138   Boethius, *Liber de Persona et Duabus Naturis*, ch. 3.
139   Locke 2009 [1690], 370.
140   Kant 2002 [1785], 46.
141   Dillon 2007, 122.

So, we can replace the question of what it takes for a person herself to recognize a reason with the question of what it takes for a rational being itself to recognize a reason. In fact, since it is *occurrent* recognition which is at issue, the question, more precisely, is what it takes for a rational being itself to occurrently recognize a reason. And presumably, such a being occurrently recognizes a reason only if the reason is immediately available to its rational capacities. And for this to be the case, the reason must be easily retrievable for use in *reasoning*. Rational capacities, after all, are capacities to put reasons to use in reasoning towards, *inter alia*, beliefs, intentions, and actions. Thus, for a rational being to occurrently recognize a reason is not only for that reason to be the content of one of her occurrent states, it is for her to be able to easily retrieve that content for use in reasoning. But this *just is* for her to be conscious of that reason. A state of consciousness, recall, is an occurrent state whose content is easily retrievable for use in reasoning. Thus, since persons are rational beings, a person herself occurrently recognizes a reason only if she is conscious of it.

And that is the personal explanation. To recapitulate: One is blameworthy only for exercises of at least the kind of control automatic doors have and rocks lack. This kind of control is the capacity to respond to reasons that are occurrently recognized. But it must be the person herself who exercises this control, not merely some part of her. So, to be blameworthy, the person herself must respond to reasons which she herself occurrently recognizes. But the person herself, being essentially and distinctively a rational being, occurrently recognizes a reason only if it is the content of an occurrent state which she can easily retrieve for use in reasoning. That is, the person herself occurrently recognizes a reason only if she is conscious of it. So, one is blameworthy only for responses to reasons which have the special distinction of being reasons of which one is conscious.

Compare this explanation to the guidance explanation. The first contrast is methodological. Levy relies on support from cognitive science for the premise that flexible responsiveness to reasons requires integration of the action's moral significance with a broad range of attitudes. The same is true of the premise that integration requires consciousness. The personal explanation, by contrast, appears to rely entirely on philosophical arguments.

As I shall now show, the personal explanation is not vulnerable to scientific falsification in the respects in which the guidance explanation is vulnerable. Recall Sripada's account, on which motivational modules allow for guidance control without the integration of the moral significance of the action with a broad range of attitudes. Sripada's account conflicts with the guidance explanation. But the personal explanation is consistent with it. Suppose Abby's motivational modules procure flexible responsiveness to reasons of which she is not conscious. On the personal explanation, Abby herself fails to recognize these reasons, and so fails to meet a necessary condition on blameworthiness.

Recall that it is also controversial in cognitive science whether integration requires consciousness. By contrast with the guidance explanation, the personal explanation can be neutral on this score. Suppose Abby's unconscious processes integrate a given set of reasons with a broad range of her other attitudes. Again, according to the personal explanation, in failing to be conscious of these reasons, Abby herself fails to recognize them. And so she fails to meet a necessary condition on blameworthiness.

So in these respects in which the guidance explanation must wait on cognitive science, the personal explanation need not. In addition, the guidance explanation has a philosophical commitment about which the personal explanation is neutral, namely the assumption that guidance control is necessary for blameworthiness. Guidance control requires that one would recognize an understandable pattern of reasons. But the control to which the personal explanation

appeals is simply the capacity to respond to reasons one recognizes. While guidance control implies this kind of control, the converse does not hold. In principle, exercising this kind of control may result simply in the recognition of *some* reasons, even if not an understandable pattern of them. Thus the personal explanation is consistent with Fischer and Ravizza's early view, on which blameworthiness-level control merely requires that one would recognize some reasons, not necessarily an understandable pattern of them.[142]

So far, we have been comparing the personal explanation to the guidance explanation. But as mentioned in §4.1, after developing the guidance explanation, Levy mentions an additional explanation in a single paragraph. Levy points out that "[r]esponsibility-level control requires both that the *agent* exercises control, and that it is *control* that the agent exercises".[143] He then maintains that for the agent to exercise control, "the agent herself…must shape behavior, and she must be receptive to a broad range of reasons".[144] Levy concludes that this requires consciousness. This explanation parallels the personal explanation in maintaining that the person herself must exercise control.

However, it depends on the same scientifically controversial theses the guidance explanation depends on. It is controversial in cognitive science whether flexible responsiveness to reasons – being "receptive to a broad range of reasons", in Levy's words – requires consciousness. For, as we have seen, it is controversial whether flexible responsiveness requires integration, and whether integration requires consciousness. And without each of these links holding, it is unclear why flexible responsiveness should require consciousness.

Even so, if sound, the personal explanation vindicates a strategic insight in Levy's additional explanation. This is the insight that, when in search for why blameworthiness should

---

[142]   Fischer and Ravizza 1991, 268-270.
[143]   Levy 2014, 114.
[144]   Ibid.

require consciousness, we do well to focus on the fact that the person herself must exercise

control. Of course, whether the personal explanation is sound depends on how well it stands up

to objections. So, to objections I now turn.

§4.6   Objections and Replies

In this section, I reply to three objections to the personal explanation.

First objection.[145] The explanation rests on the distinction between what *part* of a person

does and what the person *herself* does. For example, it is said that, while he was sleepwalking,

the part of Parks' brain which implements action scripts recognized certain visual cues, but not

he himself. But in ordinary language, we do not hesitate to say *he* recognized visual cues. Indeed,

talk of a part of his brain is awkward in comparison. So if ordinary language is our guide, it

seems often there is no important difference between what part of a person does and what the

person herself does.

Reply to first objection. Important distinctions can lurk behind ordinary language.

Andrew Bailey introduces such a distinction:

> Certain items have properties in a primary or nonderivative sense…while other
> things enjoy these properties only derivatively or by proxy. Smokey's toe is over
> the line in the primary or nonderivative sense; and so Smokey himself is over the
> line in a derivative or secondary sense. Smokey enjoys this distinction by proxy
> and only because he is related to his toe in some special way (by parthood, in this
> case).[146]

Imagine two cases. In the first, Smokey's toe is over the line, while the rest of him is not. In the

second, all of him is entirely over the line. In both cases, it is natural to say that Smokey is over

the line. But the cases are importantly different. In the first case, he qualifies as over the line

"only because" his toe, which is a part of him, is over the line. But in the second case, it is not

---

[145]   I thank Janet Levin for a discussion which inspired this objection, though I'm not sure I express it as she would.
[146]   Bailey 2015, 164.

merely because something distinct from him is over the line that Smokey qualifies as over the line. Instead, Smokey *himself* is over the line in a sense in which he himself is not over the line in the first case. To use Bailey's terminology: In the first case, Smokey has the property of being over the line *derivatively*. In the second, Smokey has this property *nonderivatively*.

This distinction allows for a response to the first objection. But first, let me offer some clarificatory remarks about it.[147] For object O to derivatively possess a property is for O to have it entirely by virtue of some distinct object's O^*'s having that same property. Thus O's nonderivative possession of a property is consistent with O's having it ultimately by virtue of more fundamental facts about other entities' having *other* properties. For example, it is consistent with O's being covered in ink in the nonderivative sense that O has this property ultimately by virtue of facts about the locational properties of atoms. Moreover, O's nonderivative possession of a property is consistent with O's having that property *partially* by virtue of facts about other entities having that same property. Indeed, in the case in which Smokey is entirely over the line, he nonderivatively possesses the property of being over the line. But he might well have this property partially by virtue of the fact that his toe is over the line. For O to nonderivatively possess a property is simply for O to possess that property while it is *not* the case that O has this property *entirely* by virtue of some distinct object O^*'s having that *same* property.

The first objection points out that, in ordinary language, we talk of how *Parks* recognized reasons while he was sleepwalking. But given the derivative/nonderivative distinction, it does not follow that he *nonderivatively* recognized reasons. Indeed, if Levy's account about the part of Parks' brain which implements action scripts is even close to correct, it looks like he has this

---

[147]   Compare with Bailey 2015, 165.

property *derivatively*. For then it looks like a part of his brain recognizes reasons in a more fundamental sense, which means he himself has this property only by inheritance. The relevant part of his brain seems to be distinct from him, and it appears to be doing all the real work.

Now, to be blameworthy, it is not enough to derivatively recognize and respond to reasons. Appropriate blame is most fundamentally directed at *the person* for what she does. So, it is first-and-foremost the person who must exercise control. And this means it is first-and-foremost the person who must have the property of responding to reasons she recognizes. That is, she must do so nonderivatively. Then we need only clarify that the premise that a person herself recognizes a reason only if conscious of it is to be read as the premise that a person *nonderivatively* recognizes a reason only if conscious of it. So, even if there is a sense in which Parks recognized reasons while sleepwalking, it is not the relevant sense.

Second objection.[148] On the personal explanation, many kinds of things recognize and respond to reasons. For example, kidneys recognize liquids and respond by purifying them. Now, kidney function is local to a part of the person. But not all unconscious processes which are reasons responsive in the relevant sense are also local. For example, if kidney function is a reasons responsive process, so is breathing during sleep. But breathing involves the entire human organism. For one, it involves the oxygenation of blood throughout the entire body.

So we should doubt the premise that a person nonderivatively recognizes a reason only if conscious of it. For in unconsciously breathing, it is first-and-foremost the person as a whole, not a part thereof, which recognizes and responds to reasons. At least, things look this way unless we are prepared to divorce persons from human organisms, as we would if, for example, we held that persons are souls.

---

[148] I thank Ralph Wedgwood for this objection.

Reply to second objection. The second objection is most forceful when we assume something like *animalism*, according to which human persons are identical with entirely physical, animal organisms of the primate species *homo sapiens*.[149] I shall first explain how animalists can reply to the objection, after which I shall explain how those who endorse other accounts of human persons can reply.

The animalist has two routes of reply. First, she can maintain that in breathing, the person only derivatively recognizes and responds to reasons. Plausibly, an animal is not identical with its *respiratory system*. And at least when one is unconscious, it is this system which appears to be doing all the work of inhalation, exhalation, oxygenation, etc. If this is right, then unconscious breathing qualifies as a human animal's recognition of and response to reasons only in a derivative sense.

One may find this reply inconclusive. One might think that, for some systems, talk of them refers not to parts of the person, but instead to *functions* the whole person can perform. Consider, for example, the metabolic or neurological systems. Perhaps these systems qualify as the person herself under different functional descriptions, similar to how one and the same person is both a doctor and a chess aficionado.

Thankfully, the animalist has another route of reply. She can offer a small amendment to the personal explanation. In particular, she can replace the premise that the person must *nonderivatively* respond to reasons she recognizes with the premise that the person must, *exercising capacities distinctive of personhood*, respond to reasons she recognizes. Indeed, it is plausible that only persons are appropriate targets of blame precisely because only persons have a certain array of capacities which must be exercised for blame to be appropriate. And since

---

[149]   Olson 2007, 23-29.

persons are distinctively rational beings, these capacities are the rational capacities to which the explanation already appeals, not capacities for respiration, metabolism, and so on. With this amendment, the question is what it takes for recognizing a reason and responding to it to be an exercise of these rational capacities. But whatever else is involved, the reason must be occurrently and immediately available to these rational capacities. And this means it must be the content of an occurrent state which one can easily retrieve for use in reasoning. That is, the person must be conscious of it.

We have seen how animalists can reply to the second objection. Those with other accounts of human persons can block the objection in a more direct way. Consider *dualism*, according to which the person is a non-physical soul which causally interacts with a physical body.[150] Even if breathing is a nonderivative process of a human organism, the person, being a soul, could qualify as breathing only in a derivative sense. So, for the dualist, the second objection does not get off the ground.

And even physicalist-friendly accounts can block the objection in this way. Consider Lynne Baker's *constitution view* of human persons.[151] On the constitution view, the human person and the human organism are two different physical objects standing in the relation of constitution. In particular, the person is constituted by the organism, in the same way that Michelangelo's *David* is constituted by a lump of marble. On Baker's underlying picture of constitution,[152] a constituted object nonderivatively has certain properties that the object which constitutes it has in a derivative sense, and vice versa. The *David* is famous nonderivatively. Derivatively, the lump of marble is famous. The lump of marble weighs over 12,000 pounds nonderivatively. Derivatively, the *David* weighs over 12,000 pounds. And so on. In any case, the

---

150   For a dualist view, see Foster 2002.
151   Baker 2007, 24-29.
152   Ibid., 23.

upshot is this: on the constitution view, it is the physical organism which nonderivatively has the property of recognizing and responding to reasons when unconsciously breathing. Only derivatively does the person have this property. So, for those who endorse the constitution view, as for the dualist, the second objection does not get off the ground.

Of course, animalism, dualism, and the constitution view do not exhaust the accounts of human persons. I mention the latter two in order to illustrate that non-animalists of quite different stripes have the resources to directly block the objection, in addition to the two strategies of reply available to the animalist.

Third objection. Persons have beliefs nonderivatively. And a belief can qualify as recognition of a reason. But a belief can be unconscious. So, contrary to the personal explanation, for a person to nonderivatively recognize a reason need not be for her to be conscious of it. Unconscious belief is a counterexample.

Reply to third objection. Recall that it is *occurrent* recognition which is required for blameworthiness. So, for the third objection to work, it must be that unconscious belief can be nonderivative occurrent recognition. Now, one might think it can. Consider an example from Levy. Emily forgets it is her wedding anniversary, yet decides "to wear the necklace her husband gave her on their last anniversary" precisely "*because* it is their anniversary".[153] One might think Emily's unconscious belief qualifies as nonderivative due to its being a belief, and occurrent due to its behavioral impact.

But, as I shall argue, any reason to think belief is nonderivative is also reason to think occurrent belief is always conscious. Our reason to think belief is nonderivative, I take it, depends jointly on the nature of beliefs and persons. For a state with proposition p as its content

---

[153] Levy 2014, 33.

to qualify as a belief in the first place, it must be that *p* is available, in some sense, to a certain array of rational capacities.[154] To illustrate, suppose that I have a state with the content *that I am over five feet tall*. Suppose also that it is impossible for my rational capacities to operate on this content, impossible, e.g., for me to infer that I am over four feet tall, or to conclude that I should buy a bed that is over five feet long, or etc. Then whatever else this state may be, it is not a belief. And the fact that the content of belief is essentially available to rational capacities explains why persons have beliefs nonderivatively. Persons are essentially and distinctively rational beings, and a state whose content is essentially available to rational capacities is first-and-foremost a state of a rational being, not a state a rational being has by proxy or inheritance.

If the content of belief is essentially available to rational capacities, belief is a nonderivative state of a rational being. But then, plausibly, a belief is *occurrent* only if its content is *immediately* available to rational capacities. For a state to be occurrent, recall, is for it to play its distinctive role. But the distinctive role of a state whose very essence is the availability of its content to rational capacities involves, at least, its immediate availability to these capacities. That is, belief so conceived is occurrent only if a state of *consciousness*, i.e., an occurrent state whose content is easily retrievable for use in reasoning. In this way, any reason we have to think belief is nonderivative is also reason to think occurrent belief is always conscious.

So, the case of Emily does not threaten the personal explanation. The occurrent state influencing her choice of necklace is not immediately available to her rational capacities. Perhaps this state is not a belief, because of the essence of a belief is the availability of its content to these capacities. Or perhaps beliefs are not essentially available in this way. But then

---

[154]   See Levin (2013, §4.3) and Wedgwood (2007, 165-167).

we have lost our reason to think beliefs are nonderivative states of rational beings. The thing to see is that under neither hypothesis does the state qualify as nonderivative occurrent recognition of a reason.

So much for the three objections and my replies.

§4.7   The Explanation of the Bridge Principle

Now we can explain The Bridge Principle. The conclusion of the personal explanation is that one is blameworthy only for responses to reasons of which one is conscious. This conclusion concerns conditions of retrospective blameworthiness, i.e., the conditions under which one is blameworthy for having done something. But it has an implication for the conditions of prospective blameworthiness, i.e., the conditions under which it is blameworthy for one to do something. In particular, it implies that it is blameworthy for one to respond to certain reasons only if one is conscious of them. This is because it is blameworthy for one to do something only if there is a possibility in which one is blameworthy for having done it. Thus, it is blameworthy for one to respond to reasons of which one is not conscious only if there is a possibility in which one is blameworthy for having responded to reasons of which one is not conscious. But that is the very thing the conclusion of the personal explanation rules out.

Now, in this dissertation, I have assumed a structural relationship between blameworthiness and blamelessness. Blamelessness is the avoiding of presently available blameworthiness, and vice versa. So, from its being blameless to respond to certain reasons in a given way, it follows that it is blameworthy to respond to them in some other way, and hence that one is conscious of them. Moreover, if one is conscious of certain reasons, it follows that these reasons are accessible to one. Thus, (I) follows from the conclusion of the personal explanation:

(I)   If it is blameless (blameworthy) for S to respond to reasons in a given way at *t*, then these reasons are accessible to S at *t*.

And from (I), all we need to arrive at The Bridge Principle is the quite plausible (II):

(II)   If (I), then whether it is blameless for S to respond to reasons in a given way at *t* is determined by the facts about which things are accessible to S at *t*.

The Bridge Principle   Whether it is blameless for S to respond to epistemic reasons in a given way at *t* is determined by the facts about which things are accessible to S at *t*.   [from (I) and (II)]

Here is why I say (II) is plausible. According to (I), the question of whether it is blameworthy, or blameless, to respond to certain reasons arises only if these reasons are accessible to one. But it is quite natural to think that whether it is blameless to respond to reasons *accessible* to one is determined by the facts about *which* reasons are accessible to one, in the first place. And it is also plausible to think that the facts about which reasons are accessible to one are determined by the facts about which things are accessible to one. If all of this is right, (II) is true.

Even so, there are two ways one might object to (II). First, might suggest that whether it is blameless to respond to reasons accessible to one is determined not only by the facts about which reasons are accessible to one, but also by the facts about what *weights* these reasons have. In reply, either the weight of a reason is essential to it, or it is not. If it is, then the facts about which reasons are accessible to one determine the facts about what weights these reasons have, and the objection falls flat. But if the weight of a reason is not essential to it, then we need only re-apply the spirit of the personal explanation. One is blameworthy for responding to something only if the person herself occurrently recognizes it, but the person herself occurrently recognizes something only if she is conscious of it. Therefore, if the weight of a given reason is something one can be blameworthy for responding to in one way rather than another, then one is conscious of that weight. Thus, it is plausible that the facts about which things are accessible to one determine not only the facts about which reasons are accessible to one, but also, if need be, the

facts about what weights these reasons have, at least in the relevant sense of 'weight'. So, (II) withstands this first objection.

Another objection one might raise to (II) runs as follows. One might suggest that no matter which items are accessible to one, it is always possible for one to have a disability which affects whether they potentially motivate one. But then, in addition to the facts about which things are accessible to one, we need to add in the facts about the subject's abilities, to fix whether it is blameworthy for her to respond to reasons in a given way.

Now, remember that we entertained this same objection in §2.5.3. There, it was aimed at

(7) The facts about which reasons can guide S are determined by the facts about which things are accessible to S.

Recall also that, in reply to this objection to (7), we considered the possibility that what it *is* for a mental state to be an accessing of something is, in part, for the subject to have the relevant abilities to be motivated by it. And notice that we argued for (II) by appeal to *access consciousness*. That is, in (II), accessibility can be read as access consciousness. But access consciousness *builds in* that the subject has the relevant abilities, otherwise the object of "consciousness" at issue is not genuinely available to her reasoning capacities. Thus, given how it is motivated, (II) is secure from this sort of objection. And barring other unforeseen objections, (II) seems highly plausible.

In this way, (I) and (II) together explain the Bridge Principle. (I) is the more substantive component of this explanation. It is also the component which follows immediately from the personal explanation.

It is worth pointing out that the personal explanation not only explains the Bridge Principle but also vindicates the connection between deontologism and accessibilism drawn by the (5)-(8) argument in §2.5.3:

(5) Rationality consists in its being blameless for S to believe.

(6) If (5), then whether believing *p* is rational for S is determined by the facts about which reasons can guide S.

(7) The facts about which reasons can guide S are determined by the facts about which things are accessible to S.

(8) So, whether believing *p* is rational for S is determined by the facts about which things are accessible to S.   [from (5), (6), and (7)]

Take (6). As we are thinking of blamelessness, it requires the present availability of blameworthiness and so, on the personal explanation, the exercise of the person's control. Now, according to the personal explanation, the person exercises control only in her responses to potentially motivating reasons she occurrently recognizes. But notice that potentially motivating reasons are one and the same as reasons that can guide one. So, blamelessness is a matter of one's response to reasons which can guide one. And this means that it is the facts about which reasons can guide one that determine blamelessness, just as (6) says. Moreover, in identifying the person's occurrent recognition of reasons that can guide one with consciousness of them, the personal explanation draws a connection to accessibility which motivates (7), especially given (II), as we have explained above.

It is nice that my explanation of the Bridge Principle vindicates the connection (5)-(8) draws between deontologism and accessibilism. For, remember, (5)-(8) encapsulates each of the main intuitive motivations on offer for accessibilism, namely BonJour's case of Norman the clairvoyant, the new evil demon, and deontologism.

Join the above explanation of the Bridge Principle with deontologism, and we have an explanation of accessibilism.

At this point, though, it is important to address the following issue. Recall from §1.2.2 that accessibilism is supposed to be a version of *internalism*. Remember also our touch-point for

internality: constancy between the normal world and the demon world. Now, on the explanation of accessibilism I have developed, does the view which is explained qualify as a version of internalism? On my explanation, rationality is determined by the facts about which things one is conscious of. Moreover, consciousness is understood as access consciousness. So, whether I have explained a version of internalism depends on whether the facts about which things one has access consciousness of are constant between the normal world and the demon world.

This last question depends on whether the contents of states of access consciousness are *narrow*, i.e., determined by facts constant between the normal world and demon world. I find it quite plausible that this kind of content is, indeed, narrow. Admittedly, there is a live debate about narrow content.[155] But in the case of access consciousness in particular, I shall offer two considerations. The first consideration is this. The content of phenomenal consciousness is surely narrow. The person in the normal world and the demon world are, at a minimum, phenomenological duplicates. But access consciousness and phenomenal consciousness appear to be connected, even if only contingently. So, there is some plausibility to the idea that access consciousness is narrow, too. And here is the second consideration. It seems the ways in which one can consciously reason are constant between the normal world and the demon world. But for this to be the case, the normal world and demon world inhabitants must be able to effortlessly and easily retrieve the same contents for use in reasoning. That is, they must have access consciousness of the same things. So, the content of access consciousness is narrow, constant between the normal world and the demon world.

Of course, these considerations are not conclusive. So, admittedly, whether I have explained a version of epistemic internalism depends on a view I have not, and cannot, defend in

---

[155] For those against narrow content, see Putnam 1975 and Williamson 2000; for those on behalf of narrow content, see Crane 1991 and Farkas 2003. In my MA thesis at CSULA, "An Internalist's Survival Guide to Twin Earth", I sided with the narrow content theorists.

any serious way in this dissertation. Thankfully, it appears to be precisely the kind of view it would *make sense* for epistemic internalism to depend on. There is a reason, after all, that advocates of narrow content are called *internalists* about mental content. It is illuminating, rather than baffling, that according to my explanation, internalism in epistemology ultimately rests on internalism in the philosophy of mind.

§4.8   Conclusion

In this chapter, I have offered an explanation of the Bridge Principle. Combine this explanation with deontologism, and we have a deontological explanation of accessibilism.

This way of explaining accessibilism raises worries, many of which I shall address in the coming chapters. Some general worries arise quite immediately. Do we really have a capacity dedicated to responding to reasons of which we are conscious? And if we do, is its exercise ubiquitous enough to explain all the cases in which, intuitively, we are blameworthy? Some appear to suggest that the answers to these questions are negative.[156] Some even appear to endorse the general approach to blameworthiness I have argued for in this chapter, and then draw the skeptical conclusions that we are never blameworthy, or at least blameworthy significantly less often than we pretheoretically might have thought.[157] In the next chapter, I address these general worries, before turning in Chapters 7-9 to a glaring worry specific to the case of belief, namely the At the Mercy of Evidence Objection.

---

[156]   See, e.g., Sher 2009.
[157]   See, e.g., Caruso 2012 and 2015.

Chapter 5: The Will

§5.1   Introduction

I offered an explanation of the Bridge Principle in Chapter 4. According to that explanation, the only thing it can be blameworthy to do is respond to reasons of which one is conscious. This explanation strongly suggests that we have a particular capacity dedicated to responding to reasons of which we are conscious, where only exercises of that capacity are blameworthy. But do we have such a capacity?

In this chapter, I argue in the affirmative, identifying this capacity with the will. I begin by looking at the what-it-is-like-ness of exercising agency, i.e., *agentive phenomenology*. Agentive phenomenology is indicative of a capacity *to experience oneself as the source of one among multiple alternatives*. This is the capacity I shall call *the will*. As we shall see, if we take the phenomenology at face value, it is plausible that the will is responsive only to reasons of which one is phenomenally conscious. At this point, I take advantage of an assumption clarified in §4.2, namely that if one is phenomenally conscious of something, then she is access conscious of it. It follows that the will is responsively only to reasons of which one is access conscious.

I should say: one can accept my explanation for the view that the only thing it can be blameworthy to do is respond to reasons of which one is conscious, and yet reject the idea that the only thing it can be blameworthy to do is exercise one's will. That is, one could get off the boat at this point, and abstain from identifying the relevant capacity with the will. I make the identification, though, for three reasons. First, as we shall see, one thing that is especially striking about the will is that it appears to be responsive only to reasons of which one is conscious. Second, as I shall argue, there is reason entirely independent of the explanation given in Chapter 4 to think only exercises of will are blameworthy. My third and final rationale concerns the

overall dialectic of the dissertation. In view of the At the Mercy of Evidence Objection we shall meet in Chapter 7, to show we have the control over belief we need to have for deontologism to be true, we need to both propose a specific account of that control as well as show that we have it. One might think that in identifying control with the will, I make things too hard on deontological explanations. But as I shall argue in Chapters 7-9, exercises of will can indeed be formings of belief. The burden can be discharged. In this way, we need not resist the intuition that blameworthiness is tied to the will, and we can gain argumentative clarity in the process.

The chapter is structured as follows. In §5.2, I examine agentive phenomenology and develop an account of the will similar to Brian O'Shaughnessy's account. In §5.3, I argue that the will, so understood, is responsive only to objects of consciousness. In §5.4, I present reason independent from the explanation in Chapter 4 to think that only exercises of will are blameworthy. After this, in §5.5, I briefly discuss Frankfurt cases, and explain why they do not raise a challenge to the account of blameworthiness proposed. I conclude in §5.6.

§5.2   Agentive Phenomenology and the Will

I shall begin by presenting and defending a picture of agentive phenomenology. I shall then use this picture to develop an account of the will, with help from Brian O'Shaughnessy.

§5.2.1 A Picture of Agentive Phenomenology

Let us begin with some examples. Consider the contrast between a memory alighting as if from the blue, on the one hand, and raising your arm, on the other. You are not conscious of any particular cause of your remembering. But in raising your arm, there is at least some sense in which it feels as if you caused your arm to go up. We might also contrast your raising your arm with your sneezing. In sneezing, you are conscious of a particular cause of what you do –

namely, an irresistible urge to sneeze. But it does not feel as if you caused your sneezing in the respect in which it feels as if you caused your arm to go up.

In contrast with remembering and sneezing, raising your arm has what Carl Ginet calls "*actish* phenomenal quality".[158] Ginet also calls it "the I-directly-make-it-happen phenomenal quality".[159] What Ginet calls actish phenomenal quality, Terry Horgan calls agentive phenomenology. Horgan offers the following description of this quality, in the case of raising your arm:

> You experience your arm, hand, and fingers as being moved by you yourself…You experience the bodily motion as generated by yourself.[160]

Moreover,

> In experiencing one's behavior as emanating from oneself as its source, one experiences oneself as being able to refrain from so behaving…[161]

Although I shall make a few comments about the details shortly, Horgan's description has a ring of truth to it. By contrast with your remembering, experiences with agentive phenomenology represent oneself as the source of what one does, in some sense. And by contrast with your sneezing, these experiences represent multiple alternatives as available, in some sense. That is, the experience involves what John Searle aptly called the "sense of alternative possibilities" that is "built right into the structure of human actions."[162] Summarizing: agentive phenomenology involves experiencing oneself as the source of one among multiple available alternatives, in some sense. That is the picture.

Now, even if this picture of agentive phenomenology is correct, one might think agentive phenomenology is *illusory*. Perhaps it is not the person herself who is the cause of what she does,

---

[158]   Ginet 1990, 13.
[159]   Ibid., 14.
[160]   Horgan 2011, 79.
[161]   Ibid., 80-81.
[162]   Searle 2001, 67.

in the relevant sense. Or perhaps, due to the truth of causal determinism, the multiple alternatives represented as available are not in fact available in the relevant sense.[163]

Given these worries, let me make a few comments about the details of Horgan's description of agentive phenomenology. First, the experience of *oneself* as source may be, more specifically, an experience of a special kind of mental state one has as the source.[164] I recommend neutrality on this front. Second, the experience of multiple alternatives available to one need not be understood as representing possibilities hostile to causal determinism. To explain, different kinds of possibility can be thought of as consistency with different kinds of facts.[165] Now, notice that physicists and ordinary folk individuate facts about people differently. Physicists are concerned with microphysical, neurological properties. But ordinary folk are concerned with people's mental states understood in terms of what it is like to live through them, what role these states tend to play in relation to other mental states, and so on. Correlative to this distinction, we can distinguish the *microphysical facts* from the *folk psychological facts*. And thus we can distinguish an alternative's being *microphysically possible* from its being *folk psychologically possible*. The first is its consistency with the microphysical facts up to the relevant time, but the second is its consistency with the folk psychological facts up to that time.[166] A number of philosophers open to causal determinism have urged that multiple alternatives are often folk psychologically possible, even if not microphysically possible.[167] And here is the upshot: the experience of the availability of multiple alternatives need not be understood as representing their microphysical possibility. It can be understood as representing

---

[163]   For arguments in this vein, see Caruso 2012, Ch. 4-7.
[164]   According to J. David Velleman (1992), this special state is one's desire to act in accord with one's reasons. Others (e.g., Mele and Moser 1994) might think it is a "proximal intention". I do not speculate on what it might be.
[165]   Logical possibility is consistency with the logical facts, metaphysical possibility is consistency with the metaphysical facts, and so on.
[166]   More precisely, the folk psychological possibility of an alternative is its consistency with the folk psychological facts up to the relevant time *plus* all the facts causally disconnected from how the person exercises her capacities at that time. For this point, see Wedgwood 2013a, 86-87.
[167]   Among this number are Anthony Kenny 1976, Ch. 8; John Searle 1984; and Ralph Wedgwood 2013a.

their folk psychological possibility. If so, what agentive phenomenology represents is entirely friendly to causal determinism.

So, we should not be too worried that the experience of the self as the source of one among multiple available alternatives is illusory. But is that kind of experience really what agentive phenomenology is like? Is our picture of agentive phenomenology accurate? Besides Horgan's and my own introspective report, there is experimental data that, if properly understood, suggests that the picture is indeed accurate.

Begin with the data. Shaun Nichols has gathered evidence that belief in indeterminism about human agency is widespread across age and culture. In one survey, subjects are told to imagine as held fixed all the initial conditions up to a particular time with respect to, on the one hand, a pot of water on a stove and, on the other, an agent making a moral choice. These initial conditions, in the latter case involving the agent's mental states and events, are described in some detail and participants are required to successfully answer comprehension questions. Most participants said that the water in the pot *had to boil*, but most *denied* that the agent making the moral choice had to make the choice he did. Most insisted he could have chosen differently. Since the sense in which the water had to boil is that it was physically causally determined, these responses indicate that most deny that human choices are physically causally determined. These kinds of results were stable across ages and cultures. This survey is just one among many with the same upshot: belief in indeterminism about human agency is widespread.[168]

Now, what explains the widespread-ness of belief in indeterminism? Nichols tries to explain it by suggesting that the masses have been duped by a certain kind of faulty reasoning.[169]

---

[168] Nichols 2012, 290-292.

[169] This reasoning runs as follows (Nicholas 2012, 298-299): The factors I can access do not determine my choice. But I have access to all the factors, at least the proximal ones, which determine my choice. So, my choice is not determined. The faulty premise is the premise that one has access to all the factors determining choice.

As a quick aside, I agree that this premise is faulty, though I should mention that we should be careful to distinguish it from the view that all of the potentially motivating reasons for which one could choose at the relevant time are accessible to one at the time. Notice that to say

But this cannot be right, because most ordinary folk have not reasoned about this issue at all, and do not so much as implicitly believe the premises of a chain of reasoning concluding with indeterminism.

Instead, it must be that participants take a description of determinism to be inconsistent with their experience of themselves as agents and then, unsurprisingly, side with their experience. But what must this experience be like, to explain this? It could not be an experience in which a mental state is experienced as *causally forcing* what one does, as in the case of sneezing due to allergies. Nor could it be simply be the absence of an experience of its cause, as in the case where I recall a memory of my childhood dog out of the blue. Deterministic processes causing events in ways we do not perceive are part of everyday life: it is easy to see, here, why what is true of a pot on a stove may be true of an unconscious mental process leading to the memory being recalled.

It could well be, though, an experience of the self as the source of the doing coupled with an experience of other doings experienced as available to one. For then it might be that the description of determinism appears to ordinary folk, whether rightly or wrongly, to cast this experience as illusory in its representation of the other doings as available to one, and so they reject determinism. So, this kind of experience could explain the widespread-ness of belief in indeterminism.

And not only this, each element of the experience plays a role in the explanation. The experience of the self as source explains why ordinary folk do not lump human agency together with deterministic processes that cause events in unperceived ways. And the experience of the other doings as available explains why ordinary folk appear disinclined to take the causation by

---

that all reasons *for which one could choose* are accessible is not to say that everything that *determines whether one chooses in a particular way* is accessible.

the self as itself deterministic. So, it is, at least, unclear how the widespread-ness of belief in indeterminism could be explained, if our picture of agentive phenomenology were not accurate. (Whether the folk's indeterminism is best interpreted as *psychological* indeterminism or *microphysical* indeterminism is another question. But perhaps many ordinary people would endorse the former and withhold on the later, if the distinction were clear to them.)

§5.2.2 An O'Shaughnessy-esque Account of Will

We have seen that agentive phenomenology involves the experience of the self as the source of one among multiple alternatives. In this section, I shall develop an account of the will on which it is the capacity to experience ourselves as the source in this way. The account builds on Brian O'Shaughnessy's account of the will.

Contemporary discussion of the will can be hard to find.[170] Many, perhaps, avoid the term 'will' in light of Gilbert Ryle's ridicule of

> …the doctrine that there exists a Faculty, immaterial Organ, or Ministry, corresponding to…the 'Will' and, accordingly, that there occur processes, or operations, corresponding to…'volitions'.[171, 172]

And those who do use 'will' often have different capacities in mind. In fact, Robert Kane has distinguished three different senses of the term in philosophy.[173] First, some use 'will' to refer to a special subset of one's desires. For example, according to Harry Frankfurt, one's will is one's effective first-order[174] desires, where a desire is effective just in case it would lead one to action in the relevant circumstance.[175] Second: some use 'will' to refer to a capacity, or set of

---

[170]   Admittedly, the phrase 'free will' is common in contemporary philosophy. But, by my lights, 'free will' is often put to use as a "semantically fused" phrase, to steal a label from Timothy Williamson (2000, 37). It often reads like jargon – perhaps for whatever may explain blameworthiness, or for our ability to choose or decide – rather than a phrase whose meaning is compositionally fixed. Rare is the discussion on which 'free will' refers to a certain capacity, *will*, in a certain condition, *free*, where the nature of this capacity and this condition are then investigated.

[171]   Ryle (2001 [1949]), 63.

[172]   In this connection, Jing Zhu (2003, 1) notes how use of the term 'volition' peters out after the 1940s.

[173]   Kane 1985, 20-21.

[174]   A desire is first-order if it is not a desire that concerns another of one's (actual or possible) desires. For example, my desire to sleep is first-order; my desire that my desire to sleep be ineffective for at least five more hours is second-order.

[175]   Frankfurt, 1971, 8.

capacities, somehow related to practical agency. For example, for Kane, will is "a set of closely related powers associated with the notion of practical reasoning."[176] In particular, he has in mind our power to form judgments about what one ought to do as well as our power to deliberate about alternative courses of action and therein form, reassess, and abandon intentions.[177, 178] Third and finally: some use 'will' to refer to what is exercised when one *strives*, *endeavors*, *attempts*, or *exerts effort* to do something. Brian O'Shaughnessy develops, at length, a picture of will in this vein.[179]

There might well be a family resemblance between effective first-order desires, practical agency, and the capacity exercised when one strives or exerts. And there might well be interesting relations between these capacities. But where is it most useful to look, if our goal is to locate the capacity such that, whenever exercised, one experiences oneself as the source of one among multiple alternatives?

We can rule out some options. Consider: A heroin addict executes an effective first-order desire to consume heroin. That is, he exercises will on Frankfurt's account. But, in doing so, he might well *fail* to be conscious of any alternative doing as available to him at the time. So, Frankfurtian will is not the capacity we are looking for.

The set of closely related powers Kane identifies as will includes our power to judge what we ought to do, our power to deliberate about alternative courses of action, and our power to form, reassess, and abandon intentions. Now, it *might* be that whenever I exercise any of these powers, I have the relevant kind of experience. But on first glance, this set of powers appears to be exercisable even when one is *asleep*, as when one awakes with an intention to do something

---

[176]   Kane 1985, 19.
[177]   Ibid., 16-19.
[178]   One might wonder whether this set of related powers are sufficiently unified to warrant the idea that to exercise any one of them is thereby to exercise one's will. For example, the power to judge that one ought to A is the power to form a *certain kind of belief*. But why think this power is part of one's will unless the power to form *any* kind of belief is part of one's will? But let us not pursue this worry.
[179]   O'Shaughnessy 1980.

one had no prior intention to do. More importantly, simply reflecting on a list of powers

somehow related to practical agency promises little insight.

But it might be that I experience myself as the source of one among multiple alternatives

exactly if I am, in some sense, *striving* or *endeavoring* to do something. On first glance, striving

seems to be the kind of thing one can do only when conscious of other things one can do at the

time. And, at least phenomenologically, nothing but the self appears to strive. This might be why

striving, at times, involves felt difficulty.

With these hints that we are looking in the right direction, let us consider

O'Shaughnessy's account of striving will in more detail. His is an account of will as a capacity to

initiate bodily actions. The first part of his picture to consider is this: to exercise will to move

one's body in a particular way requires *being aware of that way of moving one's body*. Consider

the following examples:

> *Learning the Sidekick*
> You are new to martial arts and are just learning the sidekick. Particular, hard to describe ways of moving your leg present themselves to you, which you alternatively try to execute – or, as the case may be, refine, or abandon.

> *Hesitating Before Washing*
> You walk to the sink to do the dishes but find them in disarray. Several different ways of beginning present themselves to you: subtly different ways of grabbing the big awkward pot, and so on.

These cases allow us to zero in on a feature of our experience widely present. You are not merely

aware of various ways of moving your arms and legs, but *conscious* of them. Moreover, some

but not all of these ways of moving you *realize* or instead *refrain* from realizing.

One might be tempted to dub as *plans* these ways of moving of which one is conscious.

One should be cautious here, though. Consciousness of a plan may seem to be, of its nature,

some kind of conceptual state, since 'plan' connotes a level of sophistication. But, in Learning

the Sidekick and Hesitating Before Washing, it is clear that at least many of the ways of moving

your legs and arms of which you are conscious are presented to you *nonconceptually*.[180]

Evidence of this fact is how out of place it would be to ask you to describe, in words, many of

these ways of moving your body of which you are aware. We might be better served by speaking

of *plans or quasi-plans*.

In place of 'plans or quasi-plans', I will speak of *contents of possible exercises of will*. In

addition, I intend to leave open the possibility that *moving one's body in a particular way* is

merely a *subtype* of the contents the realization of which (or refraining thereof) qualifies as an

exercise of will. That is, I shall leave open what could qualify as a content of a possible exercise

of will. For example, a mental activity or event, in principle, could qualify as the content of a

possible exercise of will.

We can see that whenever will is exercised, one experiences multiple alternatives as

available to one. With respect to each content of which one is conscious, one experiences the

alternatives of realizing and refraining from realizing. But is it true that whenever will is

exercised, one experiences oneself as the *source* of what one does?

To answer this question, let us turn to another element of O'Shaughnessy's account. An

exercise of will is an event "*so* special that…its happening *in* one is necessarily never its

happening *to* one."[181] To see what O'Shaughnessy has in mind, consider

> *Under-Described Dishwashing*
> You intend to wash the dishes after lunch, and you realize that lunch is over. You
> remember your intention. Then, conscious of various ways of moving your body,
> some of which would help bring about the dishes getting washed, you become
> aware that your body is moving in such a way that the dishes do, indeed, get
> washed.

Typically, when a description like the above is true, the event is something you have *done*, not

something with respect to which you were passive. Even so, the description does not *entail* you

---

[180]   For detailed discussion, see O'Shaughnessy 1980, Volume 1, Ch. 7.
[181]   Ibid., 260.

were active. It is consistent not only with your behavior being controlled by aliens, but also with

your occurrently knowing that it is being so controlled. It is consistent with your sitting back

passively. As O'Shaughnessy puts it, becoming aware that one's body is moving in the relevant

way "occurs in passive as well as active situations."[182]

To avoid under-description, we need a case like the following, where italics mark its

divergence from Under-Described Dishwashing:

> *Adequately Described Dishwashing*
> You intend to wash the dishes after lunch, and you realize that lunch is over. You
> remember your intention. Then, conscious of various ways of moving your body,
> some of which would help bring about the dishes getting washed, you *strive* or
> *endeavor* or *exert effort* to move your body so as to get the dishes washed.

The italics signify a very special kind of event. If one has trouble seeing the contrast between

Under-Described Dishwashing and Adequately Described Dishwashing, it might be helpful to

isolate the special kind of event in question from its usual concomitants. So, consider an example

inspired by William James:[183]

> *Paralyzed Arm*
> Unbeknownst to you, your arm is paralyzed. You are blindfolded. Your doctor
> asks you to raise your arm. After a few moments of hesitation, you obey – or, put
> carefully, do everything in your power to obey. After the blindfold is removed,
> you are surprised that your arm has not moved.

In this case, you *exert* or *strive* to raise your arm, after being conscious of a particular way of

moving it (namely: up!). You are clearly active. Moreover, you also would have been active if

you had disobeyed the doctor, and refrained from exerting to raise your arm.

Something striking about these special, active events in Adequately Described

Dishwashing and Paralyzed Arm is that one experiences oneself as their source. We do not fail to

experience what is done as caused by anything in particular. Nor do we experience it as causally

---

[182]   Ibid., 262.
[183]   Here is the passage from James 1950, 105: "Close the patient's eyes, hold his anaesthetic arm still, and tell him to raise his hand to his head; and when he opens his eyes he will be astonished to find that the movement has not taken place."

forced by some mental state we are in. Rather, in these cases, we experience ourselves as the source what we do, in the midst of other things we experience as available to us.

On the account so far developed, an exercise of will involves phenomenal consciousness of contents of possible exercises of will as well as of the self as the source of their realization. So the same holds of access consciousness, on my assumptions.

But I should mention that the view that will requires consciousness in these ways conflicts with O'Shaughnessy's position, on which a dreamless sleeper who stretches out her leg might well be exercising her will.[184] Now, I am willing to stipulate that 'will' refers to *conscious* O'Shaughnessian will. However, it is worth pointing out that the view that will requires consciousness appears to respect a genuine joint in nature.[185]

Some findings in neuroscience suggest as much. Neuroscientists have noticed that prior to observable movements the agent would classify as exercises of will, there occur "negative shifts in electrical potentials" in her brain.[186] This is called a 'readiness potential', or 'RP'. Keller and Heckhausen discovered that RPs occur both in entirely unconscious mental processes as well as in conscious ones.[187] But they also found that RPs in unconscious events and RPs in conscious events relate to different regions of the brain. In their words,

> The different scalp distributions of the two types of RP indicate that unconscious movements can be attributed to the activation of a contralateral process (lateral premotor system (LPS), primary motor cortex), whereas voluntary spontaneous motor acts seemed to be predominated by the medial premotor system (MPS).[188]

Strict O'Shaughnessian will, since sometimes unconscious, appears to involve two very different neurological systems. Conscious O'Shaughnessian will – that is, simply, *will* on my account – is

---

[184] O'Shaughnessy 1980, Volume 1, 97.
[185] There is a reason George Sher (2009, 9) says "will appears to be essentially a conscious phenomenon".
[186] This appears to have been discovered by Kornhuber and Deecke 1964.
[187] Keller and Heckhausen 1990.
[188] Ibid., 351.

associated with just one of these systems. In being restricted to consciousness, will in my sense

aligns with a real joint in nature.

§5.3   The Will's Range of Responsiveness

We are addressing the question of whether we have a capacity dedicated to responding to

reasons of which we are conscious, where only exercises of that capacity are blameworthy. In the

last section, I identified the capacity to experience oneself as the source of one among multiple

alternatives as the will. In this section, I shall argue that the will, so understood, is responsive

only to objects of consciousness. In the next section, I shall argue that there are reasons

independent of the explanation offered in Chapter 4 to think that only exercises of will are

blameworthy.

One might argue that a person is *not* conscious of each of the reasons to which the will is

responsive. For consider the literature on *priming*. Shaun Nichols relates the following case:

> In one study, participants primed with words related to rudeness were more likely
> to interrupt an experimenter than were those primed with words related to
> politeness. But none of the participants showed any awareness of the effect of the
> prime…[189]

Let us grant that the participants' interruptions are exercises of will. One might infer, then, that

the fact that they have been primed is a reason for which they exercise their will in the manner

they do. And this is to say that will, in this case, is responsive to reasons of which the agent is not

conscious, for they are not conscious that they have been primed.

But this inference from the phenomenon of priming is hasty. It is possible to exercise

one's will only when one is conscious of a content of a possible exercise of will. And

unsurprisingly, which exercises of will occur depends in part on *which* contents present

---

[189]   Nichols 2012, 303. The case is from Bargh et al 1996, 234.

themselves to one in the first place. So, if contents involving interruption present themselves much more frequently to one person as compared to another, it is unsurprising if she exercises her will so as to interrupt more frequently. Now, plausibly, priming had the effect of making contents involving interrupting the experimenters, perhaps with associated urges to realize them, occur more frequently for participants primed with words related to rudeness. Then it can only be expected that these participants interrupted more frequently.

But it does not follow that their exercises of will were *on the basis of* or *in response to* the fact that they were primed. The priming artificially induced certain temptations. But merely being tempted is not yet the *will*'s responding to anything. Instead, it is one's response to the temptation that is one's exercise of will. And we have absolutely no reason to deny that the temptation itself – namely, the contents associated with various urges or apparent reasons – is something of which the participants were conscious. Their priming was not a reason on the basis of which they exercised their will, but instead merely the causal genesis of conscious contents, urges, reasons, etc., on the basis of which they exercised their will.

This kind of explanation of priming is quite general. Moreover, it would allow us to explain, without any need for debunking, data Nichols has observed. In particular,

> (A) "…people have a kind of default presumption that the influences on decisions are introspectively available…"[190]

while, at the same time,

> (B) "…factors influencing associations and urges are regarded as less introspectively accessible than those influencing choice."[191]

Cast into my framework, (A) is the datum that people presume that exercises of will are only responsive to things of which one is conscious. The explanation of (A) to which we can help

---

[190]   Nichols 2012, 300.
[191]   Ibid.

ourselves is that people have this presumption precisely because will *is* a capacity entirely responsive to things of which one is conscious. We do not need to debunk the presumption, as Nichols attempts to do.[192] And we can explain (B) as follows. First notice that a factor "*influencing* associations and urges", like priming, is a *cause* of its being the case that the subject has before her conscious purview something for her will to respond to. We can explain (B) by saying that people are actually pretty good at dividing away things will is *not* responsive to, like the causes of associations and urges, from things the will *is* responsive to, like the conscious associations, urges, reasons, etc., themselves.

So, plausibly, will is exclusively responsive to reason of which the subject is conscious.

And there is another way to argue for this conclusion. Remember that contents of possible exercises of will are plans or quasi-plans of which the person must be *conscious*. Undoubtedly, the will is responsive to which behavior or activity is specified by these contents. Now, if one is conscious of a particular reason, it is easy to see how, similarly, the will could be responsive to the reason. For such reasons could either constitute or inform the features of a content of which one is conscious, in a manner coloring or filling in *what it would be* to realize or refrain from realizing that content. Since one is conscious of them, these reasons constitute or specify features of the content of which one is conscious, and in responding to those more fully fleshed out or colored contents, one thereby responds to those reasons. And so it is easy to see how the will could respond to such reasons.

But it is puzzling how the will could be responsive to a reason of which one is not conscious. It could not specify any features of the contents of possible exercises of will of which one is conscious. And this makes it mysterious how the will, a capacity to realize or refrain from

---

[192]   Nichols 2012, 303-305.

realizing contents *of which one is conscious*, could be responsive to such reasons, at all. So, as soon as it was clarified in §5.2 that the will is a capacity to realize contents of which one is conscious, it appears inevitable that the will is responsive only to reasons of which one is conscious.

Let us assume that the above is on track, and so that the will is only responsive to reasons of which one is conscious. This is a striking feature of the will. Recall a point I made when discussing Kane's "set of closely related powers associated with the notion of practical reasoning."[193] We saw that it is possible to exercise the power to form an intention even while entirely unconscious, as when one wakes up with an intention to do something one had not prior intention to do. It might be hard to locate another capacity essentially connected to consciousness. And recall our overall goal: to identify a capacity dedicated to responding to reasons of which one is conscious, where only exercises of that capacity are blameworthy. As I had advertised, its striking connection to consciousness is reason to take seriously the idea that the will is identical with that capacity. In the next section, I offer another reason.

§5.4   Being the Source in the Right Way

There is an old idea that one is blameworthy for something only if one is the *source* of it in the right way. First, I shall suggest that merely by reflecting on the phenomenology of the will, there is reason to think one is the source of something in the right way only if it is an exercise of will. Then I shall suggest that this argument finds further vindication in an objection J. David Velleman raises to a certain class of views about what it is to be the source in the right way.

Let us begin with some examples. Randomly, I recall a memory of my childhood dog. Then I paint a wall in my apartment orange. Not all of these things are apt for blame. I could not

---

[193]   Kane 1985, 19.

be blameworthy for sneezing, or recalling my dog. But, intuitively, I could be blameworthy for painting the wall orange. And what drives the differential treatment of these examples is not the thought that I could harm someone by painting the wall orange, while I could not harm anyone by sneezing or recalling. No: even if sneezing or recalling were to harm, I could not be blameworthy for sneezing or recalling. I could not be blameworthy for sneezing or recalling because, though these are things I have done, they do not find their source in *me* in the right way.[194]

Philosophers have found themselves under pressure to make good on this idea that one is blameworthy for only things one is the source of in the right way. Begin, for example, with compatibilists about determinism and blameworthiness in the early modern period, like Thomas Hobbes and David Hume. They held that one is the source of a doing in the right way just in case one does it because one wants to.[195] They held that this suffices to make sense of blameworthiness, but is also consistent with determinism. For example, if one steals because one wanted to steal, surely one is blameworthy for stealing. That one's stealing occurred because one wanted to steal is the doing's having one as its source in the right way, so Hobbes and Hume thought. And their view could explain why I could not be blameworthy for sneezing or recalling. Neither of those doings occurred because I wanted them to, and so neither has its source in me in the right way.

Moreover, Hobbes and Hume were wont to argue that libertarians cannot explain how one could be the source in the right way. They argued as follows. If what one does is not

---

[194] In saying that these events did not find their *source* in me, am I employing 'source' metaphorically? A source of something is where it comes from, whether ultimately or proximately. Surely the notion of 'coming from' is a causal notion. So, one might think the metaphor lies in that the source is said to be a person rather than an event. But can't one hold that the occurrence of certain mental events is simply *what it is* for a person to be a cause in the relevant way? Moreover, we need not rule out, from the start, agent causal views. So, I do not see why we must think 'source' is used metaphorically here. But even if it is a metaphor, it is nifty.

[195] As Anthony Kenny (1976, 122-123) would explain it, Hobbes and Hume hold that one is the source just in case one exercises *the liberty of spontaneity*, where one exercises this liberty just in case one does something because one wants to do it.

determined by prior mental states and events, as on libertarianism, it is simply *random*, and so does not find its source in one in the right way.[196]

Wary of this objection from randomness, early modern agent causal libertarians[197] held that causal indeterminism, even pinpointed at just the right places and times, is not enough. Such libertarians held that, in addition, *the person* must cause what she does, in a sense rising above deterministic causal relations between events. On this view, substances themselves stand in causal relations, and causation is not always deterministic. And on this view, my sneezing and my recalling do not find their source in me because, quite literally, I did not cause them. But my painting the wall does because I did cause it.

Descendants of early compatibilism and agent causal libertarianism have given increasingly sophisticated analyses of what it is for a doing to find its source in one in the right way. But what is striking is that, even in the disagreements, we see agreement on certain general constraints that must be met. First, there must be a *special way* one is part of the causal story. Notice that I am part of the causal story resulting in my sneezing, but not, for example, in the sense that I sneezed because I wanted to, or in the sense that I agent-caused my sneezing. The latter ways of being causally involved are stabs at the special way required. Second, one's doing cannot occur *randomly*, as a matter of brute chance or luck. This kind of randomness is not ruled out in the case in which I recall the memory of my childhood dog, even if my recalling, in another sense, is not random at all. Requiring that the doing is something I wanted, or that I agent-caused it, are also attempts to rule out the relevant kind of randomness. Combining the constraints we have observed, one must be part of the causal story in a *special*, *randomness-excluding* way.

---

[196]   See Balaguer 2004, 379.
[197]   Like, I suspect, Thomas Reid; for a contemporary example, see Timothy O'Connor 2000, Ch. 2.

And now we can draw a connection to the will. The special kind of involvement early compatibilists highlighted with want-caused doings and agent causal libertarians highlighted with agent-caused doings is, intuitively, absent from a case in which one experiences what one does as *causally forced* by a mental state one has. In the case of early compatibilism, perhaps this phenomenology-driven intuition is why many have been tempted to say that agents can do otherwise in *some* sense – e.g., that one can do otherwise *if one wanted to*. On the other hand, a case in which one does *not* experience what one does as caused in a particular way seems like a case in which *randomness* is afoot. What one does in cases with this experiential structure appears to be a matter of luck in the relevant sense.

By contrast, consider an exercise of will, wherein one experiences the self as the source of one among multiple available alternatives. Here one experiences the self as involved in a special manner. And while multiple alternatives are experienced as available, in experiencing the self as the source, one experiences the self as involved in a randomness-excluding manner, as well. In experiencing the self as the source of what one does, one experiences oneself as involved in a special manner. So, intuitively, an exercise of will is a case in which one is causally involved in the appropriately special, randomness-excluding way. And so, intuitively, it is a case in which one is the source in the right way. And as we had said, being the source in the right way is, plausibly, a condition on blameworthiness.

The above argument for the role of the will in blameworthiness proceeds directly on the basis of the phenomenology of the will. But this kind of argument finds further vindication in J. David Velleman's objection to what he calls *the standard story* of being the source in the right way. Here is

the standard story   For S to be the source of a doing in the right way is for it to be (non-deviantly) caused by a special class of attitudes S has – e.g., by an intention S has, or by a combination of desires and beliefs S has, or etc.

Velleman lodges the following complaint against the standard story:

> In this story, reasons cause an intention, and an intention causes bodily movements, but nobody – that is, no person – *does* anything. Psychological and physiological events take place inside a person, but the person serves merely as the arena for these events: he takes no active part.[198]

Insofar as we might be blameworthy for an event, we do not think of ourselves as mere "arenas" in which the event occurs, even if caused by an intention, desire, or belief. Instead, we think of *ourselves* as doing the deed. Or so Velleman suggests.

One might defend the standard story by pointing out that, so far from leaving out the agent, the standard story *describes her* by describing the non-deviant causal relationships between her attitudes and what she does. Nevertheless, Velleman still finds an intuitive difficulty with the standard story, which he brings to light with the following example:

> Suppose that I have a long-anticipated meeting with an old friend for the purpose of resolving some minor difference; but that as we talk, his offhand comments provoke me to raise my voice in progressively sharper replies, until we part in anger. Later reflection leads me to realize that accumulated grievances had crystallized in my mind, during the weeks before our meeting, into a resolution to sever our friendship over the matter at hand, and that this resolution is what gave the hurtful edge to my remarks.[199]

We can suppose that this is a case in which desires and beliefs caused an intention to sever friendship, and that this intention caused the severing. Moreover, we can suppose that there is no "strange perturbation" in the causal relationships involved: it is not a case of deviant causation.[200] So, the standard story implies that the agent was the source in the right way. But, according to Velleman, it is more accurate to say that the agent's desires induced the intention

---

[198]   Velleman 1992, 461.
[199]   Ibid., 464.
[200]   Ibid., 464.

than that the agent formed it. Moreover, viewing the behavior as "directly governed" by this intention is, as Velleman puts it, "precisely what leads to the thought that as my words became more shrill, it was my resentment speaking, not I."[201] And that is just to say that, intuitively, the standard story does not give us sufficient conditions for one being the source in the right way.

I'm not sure Velleman's is the clearest example of a doing's meeting the standard story's conditions while, at the same time, not being something one is blameworthy for.[202] But the example does appear to put some intuitive pressure on the idea that non-deviant causation by a special class of attitudes is sufficient for one's being the source in the right way. At the very least, we should be prepared to say more about the manner of causation, or the nature of the special class of attitudes, and so on.

The question, for our purposes, is *where* this intuitive pressure comes from. And here, the idea that one can be blameworthy only for exercises of will can help. In Velleman's example, the events constituting the severing of the friendship appear to alternate between doings that take the agent by surprise, as it were, and doings experienced as inevitable or uncontrollable at the time. On the one hand, there is no *x* such that the agent experiences *x* as the cause of the sharpness of the first reply to the first offhand comment. But, on the other, after the agent's resentment has become occurrent – in which is crystallized an intention to sever the friendship – he does not experience pulling back or responding kindly as doings available to him at the time. That is, what is absent from the case is precisely an exercise of will, i.e., an event in which one experiences one's self as the source of one among multiple available alternatives.

---

[201]   Ibid., 465.
[202]   Though see Ibid., 465, footnote 12. By my lights, it is worth asking whether the agent failed to pause, introspect, and discern whether he intended to sever the friendship – perhaps after, say, his reaction to the second "offhand comment". And perhaps if he hadn't failed in this way, he could have reassessed his intention, rather than be "directly governed" by it. However, I should admit that I think Velleman's example *could* be a case in which the relevant mental states "directly governed" the agent *all the way through*, as it were. Then the intuition remains, one might think, that the agent is not the source in the right way.

Now, Velleman admits, as he should, that the agent in the example might well be blameworthy for severing the friendship *in a sense*, because he may have negligently allowed resentment to take hold. But remember from §3.4 that the blameworthiness at issue in this dissertation involves the violating of a basic obligation to respond to one's reasons appropriately at the very time in question, and which cannot derive from one's past doings.

The overall point is that, entirely independent of my explanation in Chapter 4, there is reason to think that one can be blameworthy only for exercises of will.

§5.5   Debunking Frankfurt Cases

My approach throughout the dissertation has been to assume that one is blameworthy for doing something only if a blameless alternative was available to one at the time. In Chapter 3, we clarified that one is blameworthy precisely for one's responses to (potentially motivating) reasons. Thus, one is blameworthy only if both a blameworthy and a blameless response to one's reasons is available to one at the time.

But consider a kind of example Harry Frankfurt made famous.[203] Imagine that I am conscious of conclusive reasons to intend to buy my friend Jared a beer, such that I am blameworthy if I fail to form this intention for those reasons. But suppose that, unbeknownst to me, a chip in my brain will prevent me from forming this intention, on the condition that I reason from or appreciate these conclusive reasons in such a way I am about to form the intention. Now, with all of this held fixed, suppose that I never reason in this way, whether out of laziness, apathy, or etc. The chip does not initialize. Have I not responded in a blameworthy way to the reasons I am conscious of, in failing to form the intention to buy Jared a beer? In this sense, I

---

[203]   Frankfurt 1969.

ought to have responded to those reasons differently. But intuitively, due to the chip in my brain, I could not have responded otherwise than I did.

I cast the Frankfurt case in terms of whether I am able to respond to reasons otherwise than I in fact did. This is because what one is blameworthy for is always precisely the particular way one responds to the reasons one has. But once we zero our focus on whether, in the Frankfurt case, I have the ability to respond to my reasons otherwise than I do, it becomes clear that *I do*. Notice that it is essential to the Frankfurt case that I could have done something that would have initialized the chip, something I did not do. So, I could have done otherwise *in some sense*. As I described the case, I could have reasoned from or appreciated the reasons I had in such a way that I would be on the verge of forming the intention. What I could have done, though, clearly qualifies as a response to reasons: if forming the state is done for these reasons in normal situations, then being on the verge of forming the state is what is done for these reasons in abnormal situations beset by unbeknownst chips in one's brain. So, it is, in fact, essential to the case that I could have responded to reasons otherwise than I in fact did. Therefore, it is not a case in which I am blameworthy for how I responded to reasons despite lacking the ability to do otherwise. Instead, as Ralph Wedgwood aptly puts it, "Frankfurt-style cases *presuppose* that it is possible for the agents to think and reason differently from how they actually did."[204] I conclude that we have nothing to fear from Frankfurt's examples.

§5.6   Conclusion

In this chapter, our goal was to locate a capacity dedicated to responding to reasons of which one is conscious, where only exercises of that capacity are blameworthy. I have argued this capacity is the will. But the view that only exercises of will are blameworthy gives rise to a

---

[204]   Wedgwood 2013a, 75.

question: are exercises of will as ubiquitous as blameworthiness appears to be? I begin to address

this question in the next chapter.

Chapter 6: Activities and States

§6.1   Introduction

If my argument thus far is on track, one is only blameworthy for exercises of will. One might worry, though, whether exercises of will are ubiquitous as blameworthiness appears to be. For example, intuitively, we are often blameworthy for forming intentions and beliefs. But is it possible for the forming of a belief or intention to be an exercise of will?

This is not obvious. In fact, one might think an exercise of will is always the execution of an intention. But if that is right, then to say the forming of a particular intention or belief is an exercise of will is to say that one executed an intention to form that intention or belief. Now, even if this kind of case is conceivable, it is odd: surely most instances of forming intentions and beliefs are not executions of prior intentions to form them. And to the point: it is strongly intuitive that we are often blameworthy for forming intentions and beliefs in these normal cases, wherein one has no prior intention to form them. But, as of yet, it is not clear how *these* events of forming intentions and beliefs could be exercises of will. That is a problem for my account, thus far.

In this chapter, I advance a proposal about how to solve this problem, and then develop the proposal in light of some objections to it. According to this proposal, intention and belief are states one is in entirely by virtue of one's engaging in particular kinds of *activities*.[205] As we shall see, this proposal opens up the possibility that the formation, maintenance, revision, and etc., of intentions and beliefs can be exercises of will.

---

[205]   The phrase 'by virtue of' is to be read in its non-causal sense. For example, we might say that a figure is a triangle by virtue of its having exactly three angles, or that a wall is red by virtue of its specific surface texture. In these examples, 'by virtue of' refers to a relation of metaphysical constitution or explanation, not causation. This relation is sometimes called *grounding*, as in Paul Audi 2012. So, I shall often also say that belief is a state *grounded* in an activity. In both cases, I mean that one has this state if and only if one is engaged in the relevant activity, and that it is entirely by virtue of one's engagement in this activity that one has the state.

One objection to this proposal runs as follows. Intention and belief are enduring states, which we have even when dreamlessly sleeping. But when dreamlessly sleeping, one is not engaged in any activities, intuitively. So, it cannot be that intention and belief are grounded in activities. Now, those who hold that intention is grounded in activity have developed a general strategy for dealing with this kind of objection.[206] The basic idea is that some activities stretch across large intervals of time and are extremely easy to qualify as engaged in, and that the agent has any dispositions essential to an intention or belief by virtue of engaging in one of these activities. In this chapter, I shall contribute to this basic idea by offering a definition of the crucial term 'activity'. As we shall see, this definition shall capture the kinds of events we tend to call activities while, at the same time, vindicating the idea that it is possible to engage in an activity while dreamlessly sleeping.

Another objection to the proposal is that it does not accomplish the work it was supposed to do, viz., open up the possibility that, in some cases, the forming of an intention or belief is an exercise of will. The problem runs as follows. It is possible to exercise one's will in the relevant way without one's arm going up, as when one's arm is paralyzed. And this shows that the relevant exercise of will is not *identical* with the raising of one's arm, when one's arm is normal. Similarly, it is possible to exercise one's will in the relevant way without one's forming the intention or belief. And this shows that the relevant exercise of will is not identical with the forming of the intention or belief. We might call this *the Identity Problem.*

I reply to the Identity Problem by clarifying that there are two ways of individuating events. On one of these ways of individuating events, the relevant exercises of will can indeed be identical with the forming of a belief or intention. On the other way of individuating events,

---

[206]   Advocates of this account of intention include Ferrero 2002, Thompson 2008, and Moran and Stone 2011.

these exercises of will are not identical with the forming of a belief or intention. But I shall argue this is not a problem, for they have the kind of relationship to blameworthiness which is necessary for my explanation of the Bridge Principle to go through, so long as we appropriately shift some of the moving parts.

This chapter is structured as follows. §6.2 contains preliminaries, including clarification of the notion of intention and crystallization of the worry that standard cases of forming intention and belief cannot be exercises of will. In §6.3, I advance my proposal that intention and belief are grounded in activities. §§6.4-6.5 develops the proposal in light of the two objections discussed above. I conclude in §6.6.

§6.2   Preliminaries

In the last chapter, I clarified that the will is the capacity which, if exercised, involves experience of the self as source of one among multiple available alternatives. The will realizes, or refrains from realizing, contents of possible exercises of will. These contents can be entertained either conceptually or nonconceptually. And the will is exclusively responsive to reasons of which one is conscious at the time.

Our central question is whether an exercise of will can be the forming of an intention or belief. The issues surrounding this question become clearer once we distinguish two different approaches to what qualifies as an *intention*, in the first place. To bring out the difference between these approaches, first consider an example. Suppose that while you are busy washing the dishes, you see a spider crawling on the counter. Immediately, you are nonconceptually conscious of a way of moving your arm that would squash the spider. This way of moving your arm is a content of a possible exercise of will. Now, you neither had a prior intention to squash that particular spider, nor a prior intention to squash spiders on the counter, nor a prior intention

to avoid squashing spiders, etc. You simply had not considered the matter. Suppose that, in fact, you exert to squash the spider at time *t*, thereby squashing the spider. Here is the question: at any point in this case, do you have an intention to squash the spider?

According to the *liberal* approach to intentions, your exerting to squash the spider at *t* is your forming and beginning to execute an intention to squash the spider at *t*.[207] Advocates of the liberal approach may point out that, beginning at *t*, there is a causal propensity towards your squashing the spider. And intentions, perhaps *inter alia*, involve such a causal propensity.

According to the *restrictive* approach to intentions, though, there is no time in the above case at which you have an intention to squash the spider. On the restrictive approach, intentions satisfy the following constraints. First, they have *conceptual* contents, i.e., *plans*.[208] For example, depending on one's background theory, I may intend *to go to the grocery store today*, or *that I go to the grocery store today*.[209] Because their contents are conceptual, intentions have a level of generality, evident by the fact that there is more than one way to execute an intention, at least at some level of detail. I can get to the store by car or instead by public transit, or, at least, I can begin stepping forward with my left foot or instead with my right. Second, following Michael Bratman, intentions are *commitments* to plans, where commitment involves dispositions to act as well as "to think or reason (or, indeed, to refrain from reasoning) in characteristic ways."[210] Importantly, it is characteristic of a commitment that it can be reassessed, abandoned, or left unexecuted.

At no time in the case in which you squash the spider do you have something satisfying the restrictive approach's constraints. In exerting to squash the spider, you did not form a mental

---

[207]   See Hieronymi (2006, 56) and Mele and Moser 1994.
[208]   See Bratman 1987, 28-30.
[209]   For the first kind of background theory, see Schroeder 2011; for the second, see Wedgwood 2001.
[210]   Bratman 1987, 9.

state with conceptual content. Remember, the content was presented to you nonconceptually. Tellingly, the content you realize was not one you could have executed in more than one way. In addition, in exerting to squash the spider, you do not appear to take on a commitment. Tellingly, there is nothing in the neighborhood which you can reassess, abandon, or leave unexecuted. You might fail to squash the spider if your arm becomes temporarily paralyzed. But you do not appear to have abandoned or failed to execute a commitment should you, thereafter, simply leave the spider alone.

Remember, our central question is whether an exercise of will can be the forming of an intention or belief. On the liberal approach, we have a quick affirmative answer to this question in the case of intention. On that approach, your exerting to squash the spider at *t* is your forming and beginning to execute an intention to squash the spider at *t*.

That being said, part of what raises an intuitive problem for the view that one is only blameworthy for exercises of will is the intuition that we are often blameworthy for forming intentions *which meet the constraints placed by the restrictive account*. It is the forming of *that* kind of state which we need to show can be an exercise of will. Accordingly, for clarity, I shall leave the liberal approach behind and assume the restrictive approach, on which intentions are commitments to conceptual plans.

But then can the forming of intentions, so understood, be exercises of will? One might think so in a small, odd range of cases, but not in all of the cases in which one is blameworthy, intuitively. One might think this on the basis of the following reasoning. The only way for an exercise of will to be the forming of an intention is if that exercise of will is the execution of an intention to form that intention. But cases like these constitute a small, odd range of cases.[211]

---

[211] Compare with Hieronymi (2006, 62-63) and Wedgwood (2013a, 77-80).

And it is strongly intuitive that we are sometimes blameworthy for forming an intention, but not because, oddly, we had an intention to form that intention. So, unless we can explain otherwise, the view that we are only blameworthy for exercises of will appears to generate the implausible conclusion that one cannot be blameworthy for forming an intention one had no prior intention to form. And notice that, structurally, the same kind of objection can be stated in terms of belief. Unless we can explain otherwise, the view that we are only blameworthy for exercises of will appears to generate the implausible conclusion that one cannot be blameworthy for forming a belief one had no prior intention to form.

§6.3   The Proposal

In this section, I offer the proposal that intention and belief are states one is in entirely by virtue of one's engaging in particular kinds of *activities*. As we shall see, this proposal opens up the possibility that the formation, maintenance, revision, and etc., of beliefs and intention can be exercises of will, whether or not one had any prior intention. I shall focus on the case of intention. Often the application to belief will be straightforward. But I save for Chapter 8 more detailed discussion of how this proposal works in the case of belief.

This proposal does not come out of the blue. Reflection on examples motivates it. Consider the following example. Suppose that there is a coin on your desk. You become conscious of a plan with the following conceptual content: *to flip the coin in exactly one minute*. It is clear that your becoming committed to this plan qualifies as forming the intention to flip the coin in exactly one minute. But how do you become committed to this plan? Well, you begin to engage in a certain kind of activity which aims towards your flipping the coin in exactly one minute. To give it a name, we can call it *the activity of flipping the coin in exactly one minute*. This activity prominently involves the monitoring of time. In beginning this activity, you engage

in something which disposes you to continue until the coin is flipped at the appropriate time, though it does not *force* you to continue to this point. So, it looks like engagement in this activity involves your being disposed in the ways essential to commitment to a plan. And how *else* could you become committed to this particular plan? A minute whizzes by, after all. So, at least in this case, it appears that you have the intention entirely by virtue of engaging in the aforementioned activity. Then forming the intention is just beginning that activity.

Even if the proposal under consideration handles this case nicely, one might think it does not handle the case of intentions which concern the more distant future. According to Michael Bratman, one can have an intention to do something well *before* one does anything about it.[212] Bratman's thought can seem quite plausible if we consider intentions about the more distant future, e.g., the intention to go to the APA in a few months. It is easy to have the thought that one can have this intention well before engaging in the activity of going to the APA.

But following Luca Ferrero, perhaps the activity of going to the APA began sooner than one might have thought.[213] The initial part of this activity is preserving the feasibility of one's going, which is a flexible activity that may involve: mentally associating a blank pad on one's desk with a list of things to administrate to get to the APA, designating future mental resources to the project in a more intuitive way, keeping the relevant part of one's schedule clear, maintaining adequate funds for transportation, remaining in a location from which you could get to the APA before it is over, and monitoring the passage of time in a general way. Plausibly, one cannot intend to go to the APA without in some manner beginning the activity of preserving its feasibility. In fact, the analogy with the case of intending to flip the coin in exactly one minute is quite close.

---

[212]   Bratman 1987, 4.
[213]   Ferrero 2002, 173-203.

When we reflect on examples in this way, it is plausible to think that, in general, intentions are grounded in activities. This kind of view is suggested by a passage in Elizabeth Anscombe's *Intention*. Anscombe says the following regarding an example in which one person asks what another is doing when the other begins to walk to the kitchen:

> E.g. is there much to choose between 'She is making tea' and 'She is putting on the kettle in order to make tea' – i.e. 'She is going to make tea'? Obviously not.[214]

Richard Moran and Martin Stone point out that in saying 'She is *going* to make tea', one expresses that she has an *intention* to do something in the future.[215] So, Anscombe's point is that there is no deep or important distinction between saying that she intends to make tea and that she is making tea. One way to cash out the idea that there is no important distinction between the two is to say that the activity of making tea grounds her intention to make tea.

In any case, the proposal that states like intention and belief are grounded in activity does not come out of the blue. And it is important to see that, if this proposal is true, forming an intention or belief is the kind of thing which could be an exercise of will. For then forming an intention or belief is just the initial phase of an activity. And it is possible to experience oneself as the source of such a phase among multiple alternatives available to one. Moreover, the overall picture is made more plausible when we remember that the contents of exercises of will can be *nonconceptually* entertained. In exercising one's will to begin the activity of flipping the coin in exactly one minute, or to begin the activity of going to the APA, or to begin whatever activity grounds one's believing *p*, one need not conceptually entertain a particular plan. It is enough to have a nonconceptual sense of how to engage in the activity, which is a much less demanding bar to pass.

---

[214]   Anscombe 1969 [1957], 40.
[215]   Moran and Stone 2011, 53.

§6.4   How Activities Could Ground Enduring States

The proposal that activity grounds intention and belief faces a powerful objection. Intention and belief are enduring states, which we have even when dreamlessly sleeping. But it is natural to think that one is engaged in no activities when dreamlessly sleeping. Therefore, so runs the objection, it cannot be that intention and belief are grounded in activities. In this section, I show that clarifying what activities are, in the first place, gives us a reply to this objection.

Begin by noticing how natural it is to say that an exercise of will is an *act*. It is very natural, for example, to say that raising one's arm is an act. In fact, to better facilitate the definition of activities I shall offer, let me stipulate that 'act' refers to an exercise of will. This stipulation has a foothold in natural language. Similarly, though my definition of activities will also be stipulative, it too has a foothold in natural language.

We shall think of an act as the smallest kind, or limit case, of an activity. The reason it is more natural to call raising one's arm an act, perhaps, is because 'activity' gets most of its mileage in reference to events of longer duration. For example, walking is an activity. Walking can last for hours. Now, one's walking for hours need not involve hours of continuous acts to move one's legs in particular ways. Instead, one might become so lost in a conversation that one's walking fades far into the conscious background and continues on quite automatically. If so, substantial temporal parts of one's actual walking are not acts.

In light of this, why do we consider walking an activity? Well, notice that, at any given time, one can pay attention to one's walking and then act to move one's legs in a particular way. This observation leads to a suggestion. Perhaps an activity is an event whose parts *can* be acts.

As we develop this suggestion, it will be important to keep in mind complex activities of significant duration. Some of these activities include periods during which the agent is asleep, if we take natural language at face value. For example, if I see a person sleeping on a cot halfway

up Half Dome in Yosemite, it makes perfect sense to say, even though she is asleep, that she is *climbing to the top of the mountain*. Similarly, there is a clear sense in which, even when asleep, she is *traveling up the west coast*, *getting a PhD*, and so on. Intuitively, these temporally extended events are not simply *happening* to her, she is *doing* them. My definition of activities shall vindicate that these are indeed activities.[216]

Let *e* be a temporal part of the climb to the top of the mountain during which the climber is (dreamlessly)[217] sleeping. Is there a sense in which it is possible for *e* to be an act? Relative to the person at the time, this is not possible. For she is asleep at the time, and it is not possible to act when asleep. But in another sense, it is possible for *e* to be an act. Immediately prior to *e*, it is possible for her wake up, consider her options, and perform the act of remaining suspended on the cot instead of turning around in descent. In this sense, it is possible for *e* to be an act.

The distinction between event types and event tokens can clarify the two senses of possibility. Event *e* is a token of many different types of event.[218] Since an event token, *e* involves all the details: the person, the time, her situation, etc. Relative to all of these details, it is not possible for *e* to be an act. So, the sense in which it is possible for *e* to be an act must abstract away from these details.

Here we can put event types to use. *e* is a token of the event type *remaining suspended on the cot*. And this event type is itself a specific way for part of the event type *her climbing to the top* to occur in an interval of time. Moreover, it is possible for remaining suspended on the cot, the event type, to be an act. The point about its being possible for the climber to have woken up, and etc., illustrates that it is possible for a person to entertain and realize a content about that

---

[216] For discussion of temporally extended activities of this sort, see Ferrero 2002, 173-204.
[217] One might think one can act while dreaming. I shall henceforth talk simply of sleep, but I refer to dreamless sleep.
[218] Here I presuppose that event tokens are sufficiently coarse-grained to be tokens of multiple types. See Davidson (1963) for a view on which events are coarse-grained, and Kim (1966) for a view on which they are fine-grained.

event type, where the resulting token has actish phenomenal quality. This is no trivial feature of an event type: it distinguishes arm raising, shifting one's attention, uttering a sentence, etc., on the one hand, from the beating of one's heart, food digestion, and perceptual experience, on the other. As I shall put it, the former event types are *candidates* for being acts, while the latter are not.

We began by thinking of *e* as a temporal part of the climb to the top during which the climber is sleeping. We have learned that the sense in which it is possible for *e* to be an act is that *e* is a token of an event type which is not only a specific way for part of her climb to the top to occur in an interval of time, but also is a candidate for being an act. As we might put it, the way in which she is climbing to the top, at that time, is the kind of thing which can be an act.

The above suggests a way of thinking about what makes an event type an activity type, namely that activity types are event types whose temporal parts are candidates for being acts. But nuance is required. Many activity types have temporal parts which are too large to be candidates for being acts. No significantly large temporal part of climbing to the top, traveling up the west coast, or getting a PhD is a candidate for being an act. This is because these event types have, built in, periods in which behavior continues on automatically, or in which one is at rest. However, if we divide these event types into suitably bite-sized temporal parts, it is plausible that each of those parts is a candidate for being an act, taken individually. We saw how this could work in the case of the sleeping climber. A bite-sized, five second increment of her remaining suspended on the cot is an event type which is a candidate for being an act. Now, we need not worry about what makes a temporal part suitably bite-sized. There are arbitrarily many ways of dividing an event type into temporal parts. Surely some of them include only suitably bite-sized parts. Thus, we should say that there is *some* way of dividing an activity into temporal parts such that each of these parts is a candidate for being an act.

Gathering together what we have learned thus far, I offer the following:

definition of activity types   An event type E is an activity type if and only if there is some way of dividing E into temporal parts $e^1 \ldots e^n$ such that (i) each of $e^1 \ldots e^n$ is a specific way for part of E to occur in an interval of time and (ii) each of $e^1 \ldots e^n$ is a candidate for being an act.

My definition is stipulative, but it regiments a simple idea with a foothold in natural language. This is the idea that an activity is an event whose parts can be acts.

As promised, this definition vindicates the idea that climbing to the top of the mountain, traveling up the west coast, and getting a PhD are activities. Getting a PhD, for example, involves significant periods of reading, class-attending, writing, teaching, etc., but also involves periods of rest and diversion in other pursuits. Its being feasible to complete the PhD requires staying within certain boundary lines (maintaining certain advisor/advisee relationships, staying in the area, etc.), and the periods of rest and diversion are specific ways for part of getting a PhD to occur exactly if they keep one within these lines.[219, 220] But even during rest and diversion, it is always possible to consider the option of stepping outside of these lines and then act to remain within them. Thus, each suitably bite-sized temporal part of getting a PhD is a candidate for being an act.

And my definition is not too permissive. It categorizes clear non-activities as non-activities. Consider *digesting one's meals*, for example. Now, it is possible to divide this event type into temporal parts each of which is a candidate for being an act. For example, consider the event type *the parts of one's body moving in some way*, which is a candidate for being an act.

---

[219]   Conceiving of what it is for part of an event type to occur in this way appears to have similarities with Ferrero's approach in his manuscript "Intending, Acting, and Doing". I thank Ferrero for sharing PowerPoint slides of some of the main ideas in that manuscript.

[220]   So long as they keep one within these lines, everything the graduate student does is, at least under some description, part of her getting a PhD. An anonymous referee has suggested that this is implausible. A graduate student may get married, but getting married, so the objection goes, does not seem to be part of her getting a PhD.

In response, notice how natural it is to say that when one was getting married, one was also getting a PhD. And while the question of *how* one was getting a PhD at that time may initially seem misplaced, it has an answer. One got married in a manner which kept it feasible for one to finish the PhD. Not all ways of getting married meet this constraint. Consider, for example, getting married to someone who requires intensive, round-the-clock care.

Every five second token of digesting food is also a token of that event type. However, that event type is not a *specific* way for part of digesting one's meals to occur. The parts of one's body moving in some way is a *genus* of which digesting one's meals is a species. It is unsurprising for such genera to be candidates for being acts. But as I mean the phrase in clause (i), all specific ways for part of an event type to occur are at least as specific as that event type. So, for example, a specific way for part of digesting one's meals to occur is the event type *the dissolving of food particles by one's stomach acids*. No such event type is a candidate for being an act, in any time interval. So, my definition correctly categorizes digesting one's meals as a non-activity.

True, it is possible to divide breathing throughout one's life into temporal parts each of which is a specific way to breathe, and each of which is a candidate for being an act. Thus, my definition categorizes breathing throughout one's life as an activity. But I take this to be a virtue of my definition. Breathing appears to be active in a sense in which digesting one's meals is not. So, my definition appears to carve things at a real joint.

Remember the objection with which we began in this section: Intention and belief are enduring states, which we have even when dreamlessly sleeping. But it is natural to think that one is engaged in no activities when dreamlessly sleeping. Therefore, so runs the objection, it cannot be that intention and belief are grounded in activities.

But in light of what activities are, given the above definition, an activity can have indefinitely wide temporal extent, including intervals during which one is asleep.[221] The fact that breathing throughout one's life qualifies as an activity is a striking instance of this point. Thus, it is not the case that one is engaged in no activities when dreamlessly sleeping. In fact, even

---

[221]   It also illustrates how my definition makes engaging in multiple activities at one time extremely easy. At any given time, a normal person is engaged in a legion of activities.

during dreamless sleep, one is engaged in a legion of different activities. So, it might yet be that

intention and belief are grounded in activity. The objection fails.

§6.5   Solving the Identity Problem

So far, I have developed the proposal that belief and intention are grounded in activity. In

Chapter 8, I shall develop this proposal in more detail in the case of belief. But remember the

work this proposal is supposed to do. It is supposed to open up the possibility that, in some cases,

the forming of an intention or belief is an exercise of will. And one might think that the proposal,

even if true, does not open up this possibility.

§6.5.1 The Identity Problem

The problem runs as follows. Recall, from Chapter 5, a case I borrowed from William

James:

> *Paralyzed Arm*
> Unbeknownst to you, your arm is paralyzed. You are blindfolded. Your doctor
> asks you to raise your arm. After a few moments of hesitation, you obey – or, put
> carefully, do everything in your power to obey. After the blindfold is removed,
> you are surprised that your arm has not moved.

In this case, you exercise your will. You experience yourself as the source of one among multiple

available alternatives.

But what *is* it that you did, in Paralyzed Arm? You did not raise your arm. So, that of

which you experience yourself the source was some kind of mental event, an exerting of effort.

This effort aimed at your arm's going up, but what it aimed at did not occur, due to the paralysis.

What this shows, one might think, is that it is not possible for an exercise of will to be *identical*

with the raising of an arm. For it is possible for the first to occur without the later. But one and

the same thing cannot have different modal profiles. Thus, in a case when your arm does go up

as normal, your exercise of will is not identical with your raising your arm, but instead with something else, namely the proximate cause of your raising your arm.

And, one might think, the same reasoning shows that it is not possible for an exercise of will to be identical with the forming of an intention or belief. Surely it is possible to exercise one's will in the relevant way, yet, due to some kind of malfunction, fail to begin to engage in whatever activity grounds the relevant intention or belief. But then, so the objection runs, it is not possible for an exercise of will to be *identical* with the forming of an intention or belief. For it is possible for the first to occur without the later. But one and the same thing cannot have different modal profiles. Thus, in a case where you do form the intention or belief, as normal, your exercise of will is not identical with your forming it, but instead with something else, namely the proximate cause of your forming it. I shall call this the *Identity Problem*.

§6.5.2 Tools to Solve the Identity Problem

In this section, I lay out tools I shall use to solve the Identity Problem. The tools I have in mind consist in two distinctions. The first distinction is that between what I shall *thin* and *thick* events.

As Roberto Casati and Achille Varzi point out, there is a measure of agreement with respect to several features of events.[222] An event like a baseball game, by contrast with an object like a ball, has the following features: it happens or occurs rather than exists, it is spread out through space and time rather than something which moves through them, and it has reasonably sharp temporal boundaries despite blurry spatial boundaries.

But despite this measure of agreement, there is an important disagreement about the nature of events. Consider an oft-discussed example. Brutus assassinated Julius Caesar on March

---

[222]   Casati and Varzi 2014.

15th, 44 BCE, stabbing him to death near the Theatre of Pompey.[223] Each of the following

occurred on that fateful day: Brutus' stabbing of Caesar, and Brutus' killing of Caesar. Some

hold that the stabbing and the killing are one and the same event. Others, though, hold that these

are different – though closely related – events. This disagreement about the case follows from a

deeper, theoretical disagreement.

For example, for Jaegwon Kim, an event is a thing's exemplification of a property

throughout a certain time interval.[224] So, on this view, the event *Brutus' stabbing of Caesar* is

Brutus' exemplification of the property *stabbing Caesar* during the fateful moments of March

15th, 44 BCE. But *Brutus' killing of Caesar* is Brutus' exemplification of the property *killing

Caesar* during the fateful moments of that day. And even if these events coincide in space and

time, they are not identical, since the property *stabbing Caesar* differs from the property *killing

Caesar*.

Kim-style events are multitudinous: spread out at an object's location, there are as many

of them as there are properties the object exemplifies at that time. A Kim-style event is like one

of a multitude of sheets of paper, differing in color or aspect, but each so *thin* as to allow massive

stacks of them in a given region of spacetime. So, let us call Kim-style events *thin* events.

On other accounts, events are much thicker. For example, Donald Davidson holds that

Brutus' stabbing of Caesar is identical with Brutus' killing of Caesar.[225] In this tradition is

Jonathan Bennett's account of events.[226] Like Kim, Bennett thinks of events as property

instances at time intervals. However, on his account, a typical event is an instance of a complex

property composed of many simpler properties. And just as a singular term referring to an

---

[223] I confess that a quick search of Google, rather than a penchant for Roman history, is responsible for the details here. See, e.g., <http://www.history.com/this-day-in-history/the-ides-of-march-julius-caesar-is-murdered>. The discussion in the literature usually omits that other conspirators were involved in the assassination. For simplicity, I will omit this as well.
[224] Kim 1966, 231-232.
[225] Davidson 1963.
[226] Bennett 2002.

individual object does not by itself tell us everything about that object, a singular term referring to an instance of a complex property does not by itself tell us everything about that property instance. Like other particulars, we refer to events under guises that represent them only partially and imperfectly. So, 'Brutus' stabbing of Caesar' refers to a property instance whose full nature is a matter of what really happened to Caesar near the Theatre of Pompey on March 15th, 44 BCE. And given what fatefully occurred there, 'Brutus' stabbing of Caesar' refers to a complex property instance that includes the property *killing Caesar*, and hence refers to the same complex property to which 'Brutus' killing of Caesar' refers. Thus Brutus' stabbing of Caesar is identical with Brutus' killing of Caesar.

While thin events are multitudinous and do not purport to cut nature at any principled joint, Bennett-style events are less multitudinous and do purport to cut nature at principled joints. For this reason, I shall call Bennett-style events *thick*.

I am skeptical whether there is a fact of the matter about whether events are *really* thick or, instead, thin. It looks to me like a semantic question. Surely we should countenance both extremely detailed and less detailed property instances. And we should not forget that Kim, in the article that set off the debate about Brutus' stabbing of Caesar, states upfront that there is a sense in which Brutus' stabbing of Caesar is indeed identical with Brutus' killing of Caesar – though Kim prefers to call this the one 'happening' that occurred rather than the one 'event' that occurred.[227] Thus, the debate looks semantic. So, our ontology shall henceforth include events both thin and thick, and we shall have to keep the distinction between them in mind.

The second tool I shall make use of to solve the Identity Problem is the distinction between *shaping* events and *forming* events.

---

[227]   Kim 1966, 232, footnote 8.

Suppose that there is a machine designed to form a snowball from some surrounding snow if it receives a quarter. The machine shapes the snow into a ball-shape and, normally, the snow adheres together, at which moment a snowball has come into existence. But suppose that there is a malfunction in the snow's molecular bonding. Suppose, perhaps, that Dr. Evil is nearby with a ray gun that causes snow to behave differently than normal. In this case, the machine does what it normally does without the snow's adhering together, and so no snowball comes into existence. Though the machine has not *formed* a snowball, it has *shaped* snow ball-wise, as we might put it.

When things are working normally, if the snowball-making machine shapes snow ball-wise, it has therein formed a snowball. This is similar to the case of exercising one's will to raise one's arm. When things are working normally, if one exercises one's will to raise one's arm, one raises one's arm. Thus, an exercise of will to raise one's arm is like the shaping of events in an arm-raising manner, whether or not one raises one's arm, and so whether or not it is the forming of an event of arm raising.

In this way, we can distinguish what I shall call *shaping* events from *forming* events. No matter one's views on the causal antecedents of action, or the formings of attitudes, this is a distinction one should countenance. It latches onto a real difference that is brought to light in cases of glitch or malfunction. Whenever there is a *successful* shaping event, there is a forming event. But not all shaping events are successful.

§6.5.3 The Solution to the Identity Problem

Recall how the Identity Problem went. It is possible to exercise one's will in the relevant way, yet, due to some kind of malfunction, fail to begin to engage in whatever activity grounds the relevant intention or belief. So, it cannot be that an exercise of will is identical with the forming of an intention or belief.

There are two ways to read the Identity Problem. On the first way, the objection refers to *thick* events. So read, the objection falls flat. Considered as thick events, a successful shaping event is identical with the relevant forming event. In the same way, considered as thick events, Brutus' stabbing of Caesar is identical with Brutus' killing of Caesar. That is, in a normal case in which there is no malfunction, exercising one's will in the relevant way is a successful shaping event. One therein beings to engage in the relevant activity. So, this shaping event is identical with the relevant forming event, namely the initial phase of the activity which grounds the relevant intention or belief. If the Identity Problem refers to thick events, it has no traction.

So, the Identity Problem has punch only if read as referring to thin events. And at this point, it is important to remember *why* we endorsed the view which we expressed as the view that only exercises of will are blameworthy, in Chapter 5. We endorsed this view because *the will* seemed to be the person's control, in light of the fact that the person's control is a capacity exclusively responsive to reasons of which she is conscious. And before this, in Chapter 4, we argued that one is blameworthy for something only if the person herself exercises control, in doing it. Importantly, the kind of blameworthiness at issue here is *original* blameworthiness: that is, blameworthiness for doing something which does not derive in any way from blameworthiness for what one did in the past. And, in a particular instance of responding to reasons of which one is conscious, original blameworthiness exclusively concerns *one's response to those reasons at that time*.

So, remembering to consider the events as thin, consider a case in which one exercises one's will in the relevant way in response to a particular set of reasons of which one is conscious. Let this set of reasons be R. Suppose that there is no malfunction; so, one begins to engage in the activity which grounds the intention or belief.

The first thing to see: one's exercise of will is a response to the reasons in R. The second thing to see: one's beginning to engage in the activity it *also* a response to the reasons in R. One does not begin the activity for *no* reason. Moreover, it is highly plausible that it is the reasons in R on the basis of which, or in spite of which, one begins the activity. In cases like this, the reasons a shaping event is a response to are the very same reasons the corresponding forming event is a response to.

Now, admittedly, it is odd that there are (at least) *two* responses to R in this case. But this is just a feature of thinking of events as thin. Thin events need not carve at any principled joint. But even thin events can be more or less complex, as the associated property can be more or less complex. So, beginning the activity is more complex than merely exercising one's will in the relevant way. For, in this case, beginning the activity involves, as a part, one's exercising one's will in the relevant way. This explains why it is *not* as if, having responded to R by exercising one's will, later one has to respond to R *again*, and begin the activity. Instead, the shaping event is *a part* of the forming event. These events, since thin, are distinct. But then when a "response" to reasons is considered in terms of thin events, each such response is really a small shotgun blast of thin events standing in these sorts of part-whole relationships to each other.[228]

And here is the upshot. Our motivation is to protect the possibility that one can be originally blameworthy for forming an intention or belief. If my arguments are on track thus far, original blameworthiness exclusively concerns one's responses to the reasons of which one is conscious at the time. In the framework of thick events, some such responses are both formings of intentions or beliefs while, at the same time, identical with exercises of will. There is no problem there, and Chapters 4 and 5 is best read in terms of the framework of thick events. In the

---

[228] For a similar account, see O'Shaughnessy 1980, Volume 2, 207-214.

framework of thin events, no such response is both the forming of an intention or belief while, at the same time, identical with an exercise of will. But many such responses are both the forming of an intention or belief while, at the same time, identical with an event which has an exercise of will as a proper part – viz., many are identical with a forming event whose corresponding shaping event is an exercise of will. And since these forming and shaping events are equally responses to reasons of which one is conscious at the time, one is originally blameworthy for "both" of them. It follows that one can be originally blameworthy for forming an intention or belief. If Chapters 4 and 5 had been written in terms of thin events, rather than saying that only exercises of will can be originally blameworthy, we would have said that only exercises of will, or forming events whose shaping events are exercises of will, can be originally blameworthy. It is not surprising that we did not go this more complicated route.

In this way, the Identity Problem is solved.

§6.6   Conclusion

According to the explanation of accessibilism developed in Chapters 4 and 5, one is blameworthy only for exercises of will in response to reasons of which one is conscious. In this chapter, I developed the proposal that intention and belief are grounded in activities. This has opened up the possibility that, in the relevant sense, an exercise of will can be the forming of an intention or belief. This possibility had better be actual, if my explanation of accessibilism is to have merit. For it is strongly intuitive that we can be blameworthy for forming an intention or belief. In the coming chapters, though, I shall address an important argument to the effect that, in the case of *belief*, this possibility is not actual. In particular, I shall address the At the Mercy of Evidence Objection.

Chapter 7: That Which is at the Mercy of Evidence

§7.1   Introduction

Consider the following objection to deontologism. Our beliefs appear to be at the mercy of our evidence. But then we do not have the control over belief we need to have, for us to have epistemic obligations. So, deontologism is false. This is the objection I called the At the Mercy of Evidence Objection, in Chapter 1.

This objection to deontologism constitutes a general challenge to deontological explanations of accessibilism. Given that obligations imply corresponding abilities, any particular version of a deontological explanation will specify a kind of control over belief we need to have, to have epistemic obligations. And for the deontological explanation to work, two things need to be established. First, it must be that whether it is blameworthy to exercise this kind of control in a particular way is determined by the facts about which things are accessible to the subject. Second, it must be established that we have this kind of control over belief, in the first place. By identifying the person's control with the will in Chapters 4 and 5, which is exclusively responsive to reasons of which one is conscious, we have passed this first bar. But is it possible to pass the second? Can the forming of a belief be an exercise of will?

In Chapter 6, I developed the idea that intention and belief might be grounded activity, which would open up the possibility that the forming of an intention or belief can be an exercise of will. However, the At the Mercy of Evidence Objection threatens to close off this possibility in the case of *belief*. For the intuition that belief is at the mercy of evidence is just the intuition that it is *not* the case that, in responding to evidence, we are the source of one among multiple available alternatives. If this intuition is correct, no response to evidence is an exercise of will; *a fortiori*, no response to evidence is a part of an activity. The worry is that, by securing the link to

accessibilism in Chapters 4 and 5 by identifying the person's control with the will, we have made it implausible that we have the control over belief we need to have for us to have epistemic obligations, and so for deontologism to be true.

In principle, there are three ways to defend deontologism from the At the Mercy of Evidence Objection. One could deny that obligation requires corresponding ability or control. Or, one could argue that we do have the relevant kind of ability to respond to evidence. Or, finally, one could argue that evidential reasons do not exhaust epistemic reasons, and so that we may have epistemic obligations even if we are at the mercy of evidence. The first route, denying the 'obligation implies ability' principle, has already been closed off: given Chapters 4 and 5, obligation implies the will. In this chapter, I shall explore the second route, i.e., the idea that a response to evidence can be an exercise of will. I shall argue that a response to evidence is never an exercise of will. This will leave only the last response – i.e., of denying that all epistemic reasons are evidential.

The chapter is structured as follows. Turning to a famous source, William Alston, I lay out the At the Mercy of Evidence Objection in §7.2. In §7.3, I examine cases in which it may appear intuitive that our responses to evidence are exercises of will. In §7.4, I lay the groundwork needed to address the examples in §7.3. In particular, I distinguish two quite different kinds of doxastic states, which I shall call *credence* and *outright belief.* Moreover, I present rough outlines of accounts of each kind of state. Then, in §7.5, I return to the cases in §7.3. After pointing out that we need to be very careful to screen off exercises of will that occur immediately before our response to evidence, I argue that responses to evidence are all and only formings of particular credences. As we shall see, these points will establish that no response to evidence is ever an exercise of will: any intuition to the contrary rests on a mistake. In §7.6, I briefly discuss the general idea that our ability to respond to evidence must be on a par with our

ability to respond to practical reasons: I suggest that, in one sense, our ability is on a par, but

that, in another sense, it is not. I conclude in §7.7.

§7.2   The At the Mercy of Evidence Objection

In the literature on deontological explanations of accessibilism, it is easy to find the

objection that we do not have the ability to control our beliefs we would need to have, for the

explanation to work.[229] William Alston's memorable article on this topic is often cited.[230] To

better see the shape of the problem, I will first look at Alston and clarify the general kind of

challenge deontological explanations face. After this, I shall clarify the precise shape this

challenge takes when we assume the picture I have developed in the last two chapters.

§7.2.1 The At the Mercy of Evidence Objection, General Form

William Alston begins by arguing that we are unable to straightaway execute intentions

to believe, or to not believe, particular propositions.[231] For example, I cannot straightaway

execute an intention to believe that "the United States is still a colony of Great Britain", even

with great monetary incentive.[232] Nor can I straightaway execute an intention to cease to believe

deliverances of perception.

One might suggest that we can execute these kinds of intentions after a course of time,

perhaps interrupted by other activities.[233] But Alston argues that whatever control we have of this

kind is unreliable. Suppose that I have an intention to believe *p*. Suppose also that I search for

evidence regarding *p*, find conclusive evidence in favor, and thereby believe *p*. Nevertheless, I

did not believe *p by* executing my intention to believe *p*. Instead, the evidence I discovered

caused my belief. For my believing to be the execution of my intention, it must be the intention,

---

[229]   See, e.g., Bergmann (2006, 77-78).
[230]   Alston 1989.
[231]   In Alston's (1989, 119-121) terminology, this is to say that we lack *direct voluntary control* over belief.
[232]   Ibid., 122. Plantinga (1993, 24) makes the same point.
[233]   This is the suggestion that we have *long-range voluntary control*, even if not direct voluntary control (Alston 1989, 133-134).

not the evidence, which causes my belief. But for this to be the case, I would need to succeed in

"fighting very strong tendencies to believe when and only when something seems true".[234]

Unsurprisingly, I will only occasionally succeed, if ever. So, even if we can occasionally execute

intentions to believe over a course of time, this kind of control could only ever be unreliable.

Finally, one might suggest that we can *influence* what beliefs we have by executing

intentions to actualize states of affairs that will cause us to have different beliefs than we would

have otherwise had.[235] Alston grants that we have influence over some of our beliefs. For

example, one can execute intentions to pursue evidence with care, courage, etc. – to be

intellectually virtuous. And intellectual virtue would, in the long run, make some difference to

which beliefs one holds.

We might even have an obligation to be intellectually virtuous. But, argues Alston, no

such obligation could be *epistemic*.[236] For an obligation is epistemic only if fulfilling it places

one in a "good position to get a true belief."[237] And intellectual virtue may not be enough to

escape a poor position to get true beliefs. For example, suppose "S has lived all his life in an

isolated primitive community where everyone unhesitatingly accepts the traditions of the tribe as

authoritative."[238] Our hapless tribesman S might be admirably intellectually virtuous, yet believe

traditions that are not truth-conducive. But then an obligation to be intellectually virtuous is not

an epistemic obligation, because it is possible to fulfill it without achieving a "good position to

get a true belief."

Put summarily, here is Alston's argument. Unable to execute intentions to believe,

whether straightaway or over a course of time, we have no obligations to exercise control over

---

234   Alston 1989, 135.
235   This is what Alston (1989, 137) calls *indirect voluntary influence*.
236   Ibid., 142-152.
237   Ibid., 144.
238   Ibid., 145.

belief. But no obligation to merely influence our beliefs, perhaps by being intellectually virtuous, is an epistemic obligation. So, we do not have epistemic obligations. Hence, deontologism is false.

Alston's argument against deontologism is sometimes treated as a two-premise inference: epistemic obligations imply that we have a particular kind of control over belief; we do not have such control; so, we do not have epistemic obligations.[239] However, I believe there is a better way to think about the general challenge behind Alston's argument. This better way becomes apparent in light of two objections to Alston's argument, as it has been presented thus far.

The first objection is that we should not conclude too hastily that we can never straightaway execute an intention to believe. Granted, I cannot straightaway execute an intention to believe that the US is still a colony of Britain. But what about cases where it is unclear what is true? Can't we, at least sometimes, just decide what to believe?

Alston has a reply. To say that one "decides" to believe *p* in a case where it is unclear what is true is to use sloppy language for what is genuinely happening – viz., that *p*, at that moment, "seems to have weightier considerations in its favor".[240] And so, at that moment, I have no more power to believe otherwise than if *p* "seemed obviously and indubitably true".[241] There is a nice unity of theory in Alston's reply. His idea is that our beliefs are *at the mercy* of our evidence.[242]

The second objection comes from Richard Feldman.[243] There are examples showing that we *can* believe by executing an intention to do so. Suppose that I intend to believe the lights are

---

[239]   Ryan 2003, 48.
[240]   Alston 1989, 125.
[241]   Ibid.
[242]   It should not worry us that I do not distinguish evidence from seeming evidence. For I assume no specific account of evidence, and surely some accounts of evidence have it that how things seem to one is evidence. See §7.5.1 for a rough characterization of evidential reasons.
[243]   Feldman 2001, 81-82.

on. So, I turn on the lights and therein believe they are on, thereby executing my intention to

believe they are on.

Feldman doesn't think the counterexample provides hope for the view that we have

epistemic obligations, for the following reason:

> The worry isn't simply that we can't voluntarily control our beliefs. We can voluntarily control them, at least in some cases. The real worry is that epistemic evaluations have to do with how we respond to evidence, and we don't have voluntary control over that.[244]

Feldman paves the way for Alston to escape the problem. If we are at the mercy of our evidence,

and epistemic obligations "have to do with how we respond to evidence", then such obligations

are in bad straits indeed.

What has emerged is a more general way to think about the challenge to deontologism:

The At the Mercy of Evidence Objection, General Form

(i)   If an obligation is epistemic, it is an obligation to respond to one's evidence in a certain way.

(ii)   If one has an obligation to respond to something in a certain way, one has a certain kind of control over how one responds to it.

(iii)   A human person never has this kind of control over how she responds to evidence.

(iv)   Therefore, no human person ever has an epistemic obligation.

[from (i), (ii), and (iii)]

This is the general form of the At the Mercy of Evidence Objection. Now, let us clarify the shape

this objection takes given the picture I have developed in the last two chapters.

§7.2.2 *The At the Mercy of Evidence Objection, Specific Form*

In Chapters 4 and 5, we secured the link to accessibilism by identifying the person's

control with the will. In light of this, we know what kind of control is at issue in the At the

---

[244]   Ibid., 83.

Mercy of Evidence Objection. One has the relevant kind of obligation to respond to something in

a certain way only if one's response to it is *an exercise of will*. Therefore, we can rewrite the At

the Mercy of Evidence Objection in a more specific way, as follows:

The At the Mercy of Evidence Objection, Specific Form

(i)   If an obligation is epistemic, it is an obligation to respond to one's evidence in a
        certain way.

(ii)   If one has the relevant kind of obligation to respond to something in a certain
        way, one's response to it is an exercise of will.

(iii)   A human person's response to evidence is never an exercise of will.

(iv)   Therefore, no human person ever has an epistemic obligation.
                                                                    [from (i), (ii), and (iii)]

This is the form of the At the Mercy of Evidence Objection that shall concern me.

Now, as we have seen, I am committed to (ii). This leaves (i) and (iii). (i) seems plausible

enough. So, if (iii) is also plausible, the At the Mercy of Evidence Objection is forceful, indeed.

On first glance, (iii) appears incredibly plausible. Moreover, its plausibility is connected

to the kind of argumentation Alston offered. When I reflect on what sense it is in which I am

unable to believe the US is still a colony of Britain, or to cease to believe the deliverances of

perception, my self-report is that, at the very least, I do not experience doing otherwise as a

response available to me. Put more precisely, in these cases, my self-report is that there is no

content of a possible exercise of will that specifies my responding differently to my evidence.

Instead, I experience the evidence of which I am conscious as causally constraining me, making

entirely unavailable the possibility of believing the US is still a colony of Britain, or of ceasing to

believe the deliverances of my perceptual experience. That is to say, these are cases in which,

intuitively, my response to the evidence of which I am conscious is *not* an exercise of will.

Moreover, they seem like perfectly ordinary cases. The temptation to generalize is strong. This is

why it appears highly plausible that (iii) is right: responses to evidence are never exercises of will.

In what follows, it shall emerge that our initial impression that no response to evidence is an exercise of will is, indeed, correct. However, as we shall see in §7.3, there are cases that do not fit nicely with this initial impression. Discerning where these cases go wrong will allow us to see *why* no response to evidence can be an exercise of will.

§7.3   Might Some Responses to Evidence Be Exercises of Will?

Consider two different kinds of cases. In the first kind of case, the evidence of which I am conscious is incredibly conclusive; in the second, the evidence of which I am conscious is much more evenly balanced. As an example of the first kind of case, suppose that I am being pelted by rain. I can see rain droplets hitting the road, striking my glasses, and my clothes are soaked. Call this the *Rain Case*. As an example of the second kind of case, suppose that I am wondering whether minds are identical to brains. When I think about neuroscience's successes in identifying neural processes correlate with mental events, I feel some pressure to believe the identity thesis. But when I think about the qualitative structure of my experience and the apparent absence of this structure in neural processes, I feel comparable pressure to deny the identity thesis. I read some philosophy articles, and I may become more or less confident in one of these positions. But we all know how this goes. Conclusive evidence is not easy to find. Call this the *Philosophy Case*.

According to Alston, I cannot but believe in the Rain Case. Moreover, according to Alston, at any given point in time in the Philosophy Case, I cannot but have the belief I do, because what I believe is fixed by the evidence I am conscious of at the time, to which I cannot respond any differently than I in fact do.

Our present question is whether Alston is right, given that the ability or capacity at issue is the will. In fact, if we look closer at each case, there is reason to suspect that our response to evidence might well be an exercise of will, after all.

Let us start with the Rain Case. When discussing the kind of control we might have over belief in the Rain Case, Ralph Wedgwood makes the following suggestion:

> There are several slightly different levels of confidence that you could have in the proposition that it is raining, and it is within your power to devote more or less of your time and mental energies to thinking about the rain…[245]

Now, Wedgwood is not concerned with the question of whether one's response to evidence can be an exercise of will, in particular. However, one might think that his suggestion indicates that, in the Rain Case, one's response to evidence can indeed be an exercise of will. Paying more attention to thinking about the rain, perhaps in a skeptical or Cartesian kind of way, is something I can exercise my will to do. Moreover, whether or not I exercise my will in this way might well affect which level of confidence I have that it is raining, even if there is no chance that my confidence will fall below, say, .95. For example, if I attend to skeptical questions about the rain, my confidence may fall to .97, while if I do not give these questions a second thought, my confidence will remain at .99.

When we think about the Rain Case in this way, it does seem, at least *prima facie*, that my response to evidence can be an exercise of will. And if this is true in the Rain Case, we may be tempted to generalize this moral to other cases in which the evidence is conclusive.

Turn now to the Philosophy Case. Suppose that I am conscious of evidence that, by my lights, indicates that it is more likely than not that the identity thesis – that minds are identical to brains – is true. At the same time, having had some experience in doing philosophy, I am aware

---

[245]   Wedgwood 2013a, 79.

that a good deal of counterevidence may show up, if I keep reading articles on the identity thesis. In this kind of situation, it appears that the following kind of scenario may obtain. I experience both *believing* and *suspending* on the identity thesis as alternatives available to me at a particular time. For example, believing might be the content of an exercise of will, while suspending is presented as what I will do if I fail to realize that content. Moreover, it might well be that if I realize the content to believe, I experience myself as the source of my believing. If all of this is right, it appears that in the Philosophy Case, my forming belief in response to my evidence could be an exercise of will. Keep in mind that the suggestion is not that I can exercise my will to believe whatever I wish about the identity thesis. For example, I do not experience *denying* the identity thesis as available to me, given the evidence for it of which I am conscious. In any case, the upshot is that, in the Philosophy Case, my response to evidence can be an exercise of will. And we may be tempted to generalize this moral to other cases in which the evidence is not overwhelmingly conclusive.

So, the overall upshot, when we think about the Rain and Philosophy Cases in these ways, is that it looks like we were too hasty in thinking that no response to evidence can be an exercise of will. On the contrary, when we think about these cases in these ways, it appears that there are many cases in which our response to evidence is an exercise of will.

§7.4   Credence and Belief

We saw above that there are cases in which, at least *prima facie*, it is intuitive that our responses to evidence are exercises of will. I think these intuitions rest on mistakes. However, to see this, it is necessary, first, to distinguish two quite different kinds of doxastic states. I have in mind the distinction between levels of confidence, or *credences*, and outright doxastic states like *belief*, *suspension*, and *denial*. In this section, I give rough accounts of these two different kinds

of states. Even these rough accounts, though, will enable us to see why no response to evidence can be an exercise of will.

§7.4.1 Credence

Let us begin with levels of confidence, or credences. Suppose I tell you that I am quite confident that it will rain sometime this week, but more confident that today is Tuesday, and even more confident that 2+2=4. And suppose I tell you that though I have little confidence that bumblebees can swim, I have even less confidence – indeed none at all – that some bachelors are married. In telling you these things, I have expressed doxastic states that are *degreed*: I have the highest degree of confidence that 2+2=4, somewhat less that today is Tuesday, somewhat less that it will rain sometime this week, significantly less that bumblebees can swim, and no degree at all that some bachelors are married. It is common to map these degrees of confidence onto the unit interval (0 to 1), where 0 represents the absence of any confidence and 1 represents the highest degree of confidence. It is also common to call these degreed doxastic states *credences*.

Now, F.P. Ramsey defines credence in terms of what we are disposed to do given our desires.[246] On his view, granting we can reflect a person's desires in assignments of utility units to outcomes, to have a credence of n in p is identical with being disposed to risk up to n multiplied by m utility units on something with an outcome of m utility units if p is true and 0 utility units if p is false. Presumably, credences have something to do with utility-risking dispositions of this kind. However, I take Lina Eriksson and Alan Hajek to have argued with plausibility that credences help *explain* utility-risking dispositions, without being identical to them.[247] They have also argued that credences help explain other dispositions, including dispositions to reason in certain ways.[248]

---

[246]   Ramsey 1926.
[247]   Eriksson and Hajek 2007, 207-208.
[248]   Ibid.

But even though credence is not identical with a utility-risking disposition, we can still say a few substantive things about credence. Moreover, as our goal is to see how forming credences relate to exercises of will, we have a special interest in seeing how credences relate to phenomenal consciousness. For exercises of will essentially involve phenomenal consciousness, i.e., the experience of the self as source of one among multiple available alternatives.

To give us data for thinking about this relationship, consider L. Jonathan Cohen's account of credence.[249] According to Cohen, to have a credence of a particular degree in p is to have a disposition to "feel it true that p"[250] to a particular degree. As Cohen puts it,

> Such a feeling takes many different forms. One may feel convinced by the evidence of its being true that p, one may feel surprised to learn of an event that is evidence against its being true that p, one may feel pleased at its being true that p, and so on….You answer the question of whether you believe that p by introspecting or reporting what you are disposed to feel about the matter (even if you have never consciously thought about it before).[251]

Note what Cohen calls 'belief' here is what we are calling 'credence'. It is clear that he thinks of the state in question as coming in degrees in a way that outright doxastic states do not.[252] In any case, Cohen's picture is that credences are dispositions to have particular kinds of feelings, to different degrees, toward propositions. Moreover, these feelings are understood as conscious, phenomenological states, since they are things we can introspect and report when asked what we think about whether p is true.

When reflecting on Cohen's account, we cannot but notice that credences, *when conscious*, have a particular phenomenal quality. Moreover, it seems clear that there are cases in which we express these states by saying that we *feel* somewhat sure, or quite sure, or very sure, about p. So, though I am open to using other words to describe this phenomenal quality, we can

---

[249]   Cohen 1989.
[250]   While I use 'p' as a propositional variable in the main text throughout, this direct quotation from Cohen (1989, 368) uses 'p' as a sentential variable. Please forgive the inconsistency caused in this particular sentence.
[251]   Ibid., 368.
[252]   Ibid., 374.

go along with Cohen and describe it as feeling *p* true to a certain degree. In any case, whatever credences are, I conclude that when they are conscious, they have this kind of phenomenal quality.

There are reasons to think conscious credences have this kind of phenomenal quality above and beyond our resonation with Cohen's description. Suppose you offer me a bet on whether it will rain this week. I reflect on the matter, and this stirs up the kind of conscious feelings of which Cohen speaks. These feelings tend to dispose me, at the time, to bet in a particular way. So, the very states that obtrude in consciousness as Cohen describes also play one of the central explanatory roles credence is supposed the play, i.e., the role of explaining utility-risking dispositions.

Of course, credences are not always conscious. So what kind of mental state are they? I propose we think of them as feelings, *given* the important clarification that on this terminology, feelings can be entirely unconscious for indefinite periods of time. In particular, I propose that a credence in *p* is a feeling of *p* being true, which has a particular degree of strength, which may or may not be conscious at the time, and which helps explain certain dispositions, including the utility-risking dispositions one has concerning *p* given one's preferences. This characterization of credence is rough, and I will have some more things to say about credence in Chapter 8. But this rough outline will suffice for present purposes.

§7.4.2 Belief

When I say that I am more confident of one thing than another, I express credences. But other things we say do not express mere credences. Consider the following examples. You ask me why I just bought a lotto ticket, and ask me whether I believe it will lose. I reply by saying that I do not believe it will lose, otherwise I would not have bought it; however, I am, alas, extremely confident that it will lose. You ask me whether I believe my car will arrive safely to

my destination rather than explode on the freeway. I say that I believe my car is safe, though

confessing that I have more confidence that my lotto ticket will lose than that my car is safe. In

saying I believe my car is safe while not believing that my ticket will lose, I express doxastic

states that differ from specific credences. I express a kind of belief that is *outright* rather than

degreed. One way to put it is that I *rely* on my car being safe in my reasoning and action, but do

not rely on the ticket losing in my reasoning and action. When we believe *p* in this outright

sense, we use *p*, at least normally, as a starting point from which to reason and act, even if we do

not have a credence of 1 in *p*.

Indeed, it appears that it is crucial that we have these kinds of outright doxastic states. For

if all of our reasoning were driven by our credences, then even very simple practical problems

would be completely intractable. Consider an example.[253] To illustrate, suppose that I am

deliberating about whether to pick up a movie from my local Redbox or instead go to the theatre

downtown. It quickly becomes clear that my decision problem hinges on two issues: how likely

it is that there will be bad traffic in getting downtown, and how much I prefer seeing Mad Max in

theatre to an oldie from Redbox at home. Suppose that I decide to go, thinking that Mad Max is

worth the non-negligible risk of traffic, even if not the certainty of traffic. In this case, the

specific credence I have that there will be bad traffic motivates my reasoning, in a rather obvious

way. But notice other credences I have that do *not* affect my reasoning. For example, my

credence that my car is safe, though high, is not 1. However, I spend no time factoring in the

possibility that my car will explode on the way. But it is not as if I have *no* doxastic attitude

towards my car's safety. Instead, it is clear that, in addition to my non-maximal credence that my

car is safe, I *believe*, outright, that it is safe. That is, I use the proposition that it is safe as a

---

[253] For similar examples, see Ross and Schroeder (2014, 262-268) and Wedgwood (2012, 323-324).

starting point from which to reason. As a little reflection shows, if we did not have these kinds of starting points from which to reason, even very simple practical problems would be, quite immediately, cognitively intractable. We are not, and it is not psychologically realistic that we become, both efficient and ideal probabilistic reasoners. To get anywhere, we need the heuristic that is outright belief.

Just as I gave a rough characterization of credence above, I should like to do so with respect to outright belief, as well. Henceforth, I shall call outright belief simply *belief*.

Consider an account of belief due to Jacob Ross and Mark Schroeder: to believe *p* is to have defeasible dispositions to treat *p* as true in one's reasoning.[254] To treat *p* as true in one's reasoning is to rely on *p* in one's reasoning, which involves not taking the possibility of *p*'s negation into account in one's reasoning. For example, when I reason about how many transportation devices I have, I do not take into account the possibility that my bike was stolen last night. Nor do I take that possibility into account when I reason about how to get to school. To be *defeasibly* disposed to treat *p* as true is to be such that there are certain abnormal or extreme conditions in which my disposition to treat *p* as true will be blocked. If a demon offers me $1 if I have a bike against eternal damnation if I do not, I will not take the bet, precisely because this would constitute a situation in which my disposition to treat *p* as true is blocked. Interestingly, Ross and Schroeder's is an account on which it falls out that belief is not reducible to credence. This is because a credence of *less* than 1 is not a disposition of any kind to treat *p* as true. On the other hand, a credence of 1 is not a *defeasible* disposition to treat *p* as true; instead, it is a disposition on which one would take the demon's bet. So, belief is no mere credence of any degree.

---

[254] Ross and Schroeder 2014, 267-268.

In Chapter 8, I shall develop my own proposal about the nature of belief. I shall propose that belief in *p* is the state one is in by virtue of engaging in the activity of attending only to possibilities consistent with *p* in one's reasoning. The dispositions grounded in this activity are *very like* a defeasible disposition to treat a proposition as true in one's reasoning. In any case, for present purposes, I shall accept the following rough characterization of belief: it is something like a defeasible disposition to treat a proposition as true in one's reasoning.

§7.5   Why No Response to Evidence Can Be an Exercise of Will

We are now in a position to diagnose the cases, discussed in §7.3, which *prima facie* appear to be cases in which one's response to evidence is an exercise of will. Recall, there were two cases, the Rain Case and the Philosophy Case. In the Rain Case, I am being pelted by rain. It was suggested that it is possible for me to exercise my will to wax philosophical about the rain, which would reduce my credence to lower than it would otherwise have been. So, it was suggested, my response to evidence in the Rain Case can be an exercise of will, after all. *Mutatis mutandis*, on this line of thought, for other cases in which the evidence appears conclusive. In the Philosophy Case, I am wondering whether minds are identical to brains. I am conscious of pieces of evidence that, by my lights, indicates the identity thesis is more likely than not. It was suggested that it is possible for me to exercise my will to either believe the identity thesis or instead suspend belief regarding it. So, it was suggested, my response to evidence in the Philosophy Case can be an exercise of will, after all. *Mutatis mutandis* for other cases in which the evidence does not appear conclusive. I have been saying that mistakes undergird these suggestions. Now we are ready to see what these mistakes are.

§7.5.1 Diagnosing the Rain Case

The suggestion about the Rain Case invoked our capacity to direct our attention. A few remarks on attention are in order. Sometimes shifts in attention seize us like foreign invaders, or at least arrive without being called for. For example, when a lightning bolt strikes thirty feet away from a person, her attention is immediately and instinctively taken over by the lightning, at least for several seconds. Moreover, sometimes a thought occurs, or recurs, without being called for. In the literature on attention, such shifts in attention are known as *involuntary*. But other shifts in attention, the so-called *voluntary* ones, are different.[255] I may explicitly focus in on a particular question, and therein ignore other things I experience myself as able to attend to at the time. In cases like these, one experiences oneself as the source of how one attends, and experiences other ways of attending as available to one at the time. That is, these so-called voluntary shifts of attention are exercises of will. I shall focus on these kinds of shifts of attention, hereafter.

Again, the suggestion about the Rain Case is that I can exercise my will to shift my attention, which will affect whether my credence that it is raining is .97 or instead .99. Now, the question we need to ask is whether my exercise of will is a response to my *evidential* reasons, or, instead, a response to my *practical* reasons.

Since that is the question, let me briefly comment on the distinction between evidential and practical reasons. I shall assume no substantive account about either kind of reason. However, I gather that we are generally quite good at telling the difference between these two kinds of reasons. In at least some sense of 'probability', an evidential reason for or against having a particular doxastic attitude toward a proposition, *p*, is something that affects (or, is

---

[255]   Voluntary attention is also called endogenous or top-down; involuntary attention is also called exogenous or bottom-up. For this distinction, see Ruff (2011, 5), Mole (2011, 67), Armstrong (2011, 87-88), and Wu (2011, 107).

taken to or experienced as affecting) the probability of p in some way.[256] By contrast, practical reasons do not affect the probability of the proposition in question. Instead, a practical reason to do something is the kind of reason that might tempt one to offer the following analysis: a practical reason to do something raises its value, either in the sense that it is a more valuable thing to do given the reason, or in the sense that the subject takes or experiences it as a more valuable thing to do given the reason.[257] I do not assume this analysis, of course. But it is indicative of the flavor of practical reasons as against evidential reasons. Henceforth, I rely on our ability to distinguish these two kinds of reasons.

Is my exercise of will to wax philosophical in the Rain Case a response to evidential reasons, or instead to practical reasons? Well, responding to a reason is doing something on its basis, or doing something in spite of it. With this clear, it is clear that waxing philosophical is not a response to evidence, after all. Shifting my attention in a particular way is one thing, and forming a particular credence is something very different, even if the first leads to the second. The point can be phrased in terms of my rough characterization of credence. When I shift my attention to Cartesian musings, I may after a course of time come to have a feeling – with a degree of .97 rather than .99 strength – of its being true that it is raining, and which explains why I tend to have the relevant utility-risking disposition, and etc. This is because I may be struck by some new piece of philosophical evidence, or come to appreciate some old evidence in a new way. My being struck in this way is my feelings about its raining changing slightly, as described. Put differently, it is my credence changing. But it is clear that my waxing philosophical is one thing, and my credence changing is something very different. And it is only the latter that is a response to evidence.

---

[256]   Definitions of evidence in terms of probability are legion. For one example, consider Williamson 2000, 186-187.
[257]   This account of practical reasons is similar to one developed in Wedgwood 2009.

If I wax philosophical for reasons of which I am conscious, these reasons are reasons that make it attractive for me to wax philosophical. Perhaps my reason is that I like philosophizing, or perhaps it is that waxing philosophical will prevent me from thinking about things I'd rather ignore. Perhaps my reason is that reducing my credence that it is raining has a particular kind of value. Even in the last case, the reason does not itself affect the probability that it is raining, and so it is not an evidential reason. So, paying attention to skeptical questions about the rain, even if this affects my credence, is not a response to evidential reasons, but instead to practical reasons.

So far, the moral is that, when looking for responses to evidence that may qualify as exercises of will, we need to screen off responses to practical reasons that occur before our response to evidence. These prior responses may be exercises of will, but that is not to the point.

Responses to evidence are all and only formings of particular credences. But no forming of a particular credence can be an exercise of will. But, in addition, we can see why the suggestion centered on an exercise of will to shift one's attention, rather than on an exercise of will to form a particular credence *directly*. This is because, once we look to see, it is clear that no forming of a particular credence can be an exercise of will. Now, for all I have argued, having a credence over a period of time might well be constituted by one's undergoing a certain kind of mental event over that period of time. But no part of this event could be an exercise of will. Remember, a credence in *p* is a feeling of *p* being true, which has a particular degree of strength, which may or may not be conscious at the time, and which helps explain the utility-risking dispositions one has concerning *p* given one's preferences. But it is not possible to experience oneself as the source of having a definite feeling of a particular strength in *p*'s truth, in the midst of other alternatives experienced as available to one. In the same way, it is not possible to experience oneself as the cause of having a perceptual experience of black text on white background rather than green text on purple background, or vice versa. Exercises of will are

ensconced in phenomenal feelings, but the *content* of a possible exercise of will – the thing presented as available for one to *do* or *refrain* from doing – is never a feeling. Feelings are essentially passive. So, what part of an event grounding a credence could be an exercise of will? The answer: none.

So, it is not possible for forming a credence to be an exercise of will. Thus, if we are looking for responses to evidence that are exercises of will, we shall have to look for responses that are not formings of particular credences. This was the suggestion with respect to the Philosophy Case, to which we now turn.

§7.5.2 Diagnosing the Philosophy Case

In the Philosophy Case, I am wondering whether minds are identical to brains. I am conscious of evidential reasons in response to which I have a credence of, say, .6 in the identity thesis. It was suggested that it is possible for me to exercise my will to either believe the identity thesis or instead suspend belief regarding it. So, according to the line of thought, it is possible for a response to evidence to be an exercise of will.

Here *belief* and *suspension* are outright doxastic states, in contrast to credence. In later chapters, I shall defend the idea that I can exercise my will to either believe or suspend in cases like these. However, it does not follow that, in cases like these, one's response to evidence is an exercise of will. The line of thought makes a jump to its conclusion.

In light of my argument in the last section, I cannot exercise my will to form a particular credence in the identity thesis. In this sense, I am stuck with the credence I have, a credence of .6. Therefore, if I am capable of exercising will to either believe or suspend, belief and suspension cannot be understood in terms of credence. But then, as I shall argue, insofar as forming a doxastic state is not a matter of forming or adjusting a credence, it is, by the same token, *not* a response to evidential reasons but instead to practical reasons. Not only is forming a

credence a response to evidence, one's response to evidence is *complete* once the credence is formed.

To see this, consider the rough characterization of belief offered in §7.4: belief is something like a defeasible disposition to treat a proposition as true in one's reasoning. Having a credence of .6 in the identity thesis is not in itself to have a belief, in this sense. But suppose that I come to a point where I am experientially faced with the alternatives of believing the identity thesis, while having a credence of .6 throughout, or rather suspending, while having a credence of .6 throughout. What sort of reasons could I base my actualizing one of these alternatives rather than the other? Whatever else we say, it is apparent that *evidential* reasons are, in this respect, quite impotent. Remember, evidential reasons to believe *p* raise (or are taken or experienced as raising) the probability of *p*. Responding to evidential reasons, therefore, affects one's *credence*. But the alternatives we are considering are alternatives in which one's credence remains unaltered. Therefore, the reasons for which I exercise my will to either believe or suspend, given that my credence is held fixed, could only be non-evidential, practical reasons. As I had said: one's response to evidence is complete once the credence is formed.

So, contrary the suggestion in §7.3, the Philosophy Case is not a case in which it is possible for one's response to evidence to be an exercise of will. Furthermore, in light of this section and the last, we are in a position to explain, in a straightforward way, why no response to evidence can be an exercise of will. Responses to evidence are all and only formings of particular credences. But no forming of a particular credence can be an exercise of will. Therefore, no response to evidence can be an exercise of will.

§7.6   Are Evidential and Practical Reasons on a Par?

Some have held that our ability to respond to evidence is on a par with our ability to respond to practical reasons.[258] If this is right in general, then we should expect that if a response to practical reasons can be an exercise of will, then a response to evidential reasons can be an exercise of will, too. In Chapter 6, I suggested that it is possible for the forming of an intention in response to practical reasons to be an exercise of will. But in this chapter, I have argued that no response to evidential reasons is an exercise of will. So, my arguments imply that our ability to respond to evidential and practical reasons is *not* on a par, at least in general.

I do not think this kind of disparity between evidential and practical reasons is, in itself, a bad thing. They are, after all, very different kinds of reasons. Even so, I should like to suggest that our ability to respond to evidential and practical reasons is on a par in an important domain or region of our mental economy. Just as formings of credences are responses to evidence, so formings of desires or preferences are responses to practical reasons. And when it comes to our ability to form these states in response to reasons, credences are to be compared to desires or preferences, not intentions. Desires or preferences, by my lights, have a functional profile very similar to that of credences. When conscious, desires or preferences involve a certain kind of phenomenal feeling of attraction to a course of action, of various degrees of strength, and they explain some of our dispositions. And just as no forming of a credence is an exercise of will, no forming of a desire or preference is an exercise of will, either.[259]

For each of these different kinds of degreed states, forming it in response to a set of reasons is experienced as one's being struck by these reasons, rather than these reasons posing to one a fork in one's path. And this might have something to do with the fact that they are degreed.

---

[258]   Ryan 2003, 63.
[259]   Compare Hansson and Grüne-Yanoff 2011, section 5.

Because these states are degreed, evidential reasons for and against *p* and practical reasons for and against desiring a particular state of affairs do not compete in such a way that one's credence or desire has to arbitrate between them. Evidence against a probable proposition *p* does not compete with the evidence for *p*: being struck by evidence against *p* is not abandoning one's appreciation of the evidence for *p*, it is just having an updated credence in *p*. The shift in credence can be quite small. And the same comments hold with respect to desires or preferences and practical reasons. Reasons with respect to these states do not offer forks in our paths, but instead just push these states around.

Therefore, desires and credences are to be contrasted with intentions and beliefs. Assuming my arguments in the coming chapters are on track, what is distinctive of both intention and belief is that it is possible for the forming of the attitude to be an exercise of will.

§7.7   Conclusion

Let us look again at

The At the Mercy of Evidence Objection, Specific Form

(i)   If an obligation is epistemic, it is an obligation to respond to one's evidence in a certain way.

(ii)   If one has the relevant kind of obligation to respond to something in a certain way, one's response to it is an exercise of will.

(iii)   A human person's response to evidence is never an exercise of will.

(iv)   Therefore, no human person ever has an epistemic obligation.
[from (i), (ii), and (iii)]

As I have said, I am committed to (ii). And while there is *prima facie* reason to doubt (iii), these reasons turn on mistakes. In fact, we have seen why (iii) is true: responses to evidence are all and only formings of particular credences, and no forming of a particular credence can be an exercise

of will. Therefore, if my argument is on track thus far, denying (i) is our only escape route from

the At the Mercy of Evidence Objection.

Chapter 8: Belief and Activity

§8.1   Introduction

The last chapter concluded that, to escape the At the Mercy of Evidence Objection, we need to reject the premise that if an obligation is epistemic, it is an obligation to respond to one's evidence in a certain way. Put differently, this is the premise that all epistemic reasons are evidential. Now, notice that even if this premise were successfully resisted, the victory would be merely Pyrrhic unless we could also establish that forming a belief can be an exercise of will. For even if some epistemic reasons are practical, my deontological explanation fails if forming a belief can never be an exercise of will.

My strategy is to first explain how forming a belief can be an exercise of will in this chapter, and then reply to the premise that all epistemic reasons are evidential in Chapter 9. This is my strategy for the following reason. As we shall see, what makes a reason epistemic is that it is the "right kind" of reason for belief. But what makes a reason the right kind of reason for belief depends on the nature of belief. And what helps us to see how forming a belief can be an exercise of will is to get clearer on the nature of belief. Once we are clearer on the nature of belief, we shall see how forming a belief can be an exercise of will. And we shall also be better situated to see whether all epistemic reasons are evidential.

In Chapter 6, I suggested that intention and belief are grounded in activity, which would explain why forming one of these states can be an exercise of will. In this chapter, I make good on this suggestion in the case of belief. I develop an account on which belief in p is the state one is in by virtue of engaging in an activity of organizing one's attention in a particular way. I shall call this activity *organizing one's attention as if p is true*. I shall argue that this account meets several constraints an account of belief should meet. But, in addition, I shall argue that this

account can explain features of belief it is unclear rival accounts can explain. In particular, my account can resolve a tension in our conception of belief, in the process vindicating ordinary talk and blaming practices that suggest we often choose our beliefs. In addition, my account can explain, in a simple and elegant way, why it can be perfectly coherent to believe *p* without having maximal credence in *p*.

The chapter is structured as follows. §8.2 lays out constraints an account of belief ought to meet. In §8.3, I develop my account of belief, and explain how it meets these constraints. In §8.4, I argue this account explains features of belief it is unclear rival accounts can explain. I conclude in §8.5.

§8.2   Constraints on an Account of Belief

An account of belief should have the resources to explain certain facts about belief. In this section, I lay out five such constraints on an account of belief.[260]

The first constraint is that belief is an enduring mental state. As Matthew Boyle says,

> We retain our beliefs even in dreamless sleep, when – on the usual understanding of "doing", at least – we are not doing anything.[261]

Beliefs are states which endure through dreamless sleep.

The second constraint concerns the core role of belief in our mental economy, and warrants more discussion. This is the fact that to believe *p* is, in some sense, to rely on *p* as *true*. The idea is expressed in different ways. According to Robert Stalnaker,

> …to believe [*p*] is to be disposed to act in ways that would tend to satisfy one's desires, whatever they are, in a world in which *p* (together with one's other beliefs) were true.[262]

---

[260]   The literature contains a number of constraints which concern how belief relates to our normative concepts. Consider the following: a belief is correct just in case it is true, a belief is rational only if one has sufficient evidence for it, and it is irrational to believe inconsistent propositions. For discussion of these constraints, see Ross and Schroeder 2014, Wedgwood 2012, and Dallmann 2014. I shall focus on constraints which do not explicitly refer to the normative.

[261]   Boyle 2011, 6.

[262]   Stalnaker 1984, 15.

Stalnaker's definition is in terms of how one is disposed to act. But there is a growing tendency to understand belief in terms of how one is disposed to reason.[263] So, Brian Weatherson says that "to believe that p is to treat p as true for the purposes of practical reasoning".[264] In similar fashion, Fantl and McGrath say that to believe p is to be "prepared to put p to work as a basis for what you do, believe, etc.".[265] Or, as Ralph Wedgwood puts it,

> If you have an outright belief in p, you will simply take p for granted, treating p as a starting point for further reasoning…[266]

I follow this trend and shall understand reliance on p as true in terms of reasoning.[267, 268]

Let us look at a kind of case often used to illustrate this core role of belief.[269] I am deliberating about whether to watch Netflix or instead Mad Max in the theatre. Watching Netflix would be fine. But so long as traffic would not delay me more than 30 minutes, I would prefer to go to the theatre and see Mad Max. I have a credence of .6 that traffic would not delay me more than 30 minutes. Incidentally, I also have a credence of .99 that my car would not break down on the way to the theatre.

There are two variations of the case. In both variations, I factor into my reasoning the possibility that the traffic would delay me more than 30 minutes. In Variation One, I do not factor into my reasoning the possibility that my car would break down on the way. That possibility may occur to me. But if it does, I ignore it. We can represent the possibilities I factor into my reasoning in Variation One with a matrix, whose rows represent courses of action, and whose columns represent states of nature:

---

[263]   In this connection, see Ross and Schroeder 2014, 263-268.
[264]   Weatherson 2005, 421.
[265]   Fantl and McGrath 2010, 143.
[266]   Wedgwood 2012, 312.
[267]   Plausibly, relying on p as true in reasoning will make us tend to rely on p as true in how we act in general. If this is right, then something like Stalnaker's account could be explained by an account in terms of reasoning.
[268]   Reasoning is commonly divided into practical and theoretical. Some hold that belief is to be understood only in terms of practical reasoning; see, e.g., Wedgwood (2012, 321). Others think belief can be understood in terms of both kinds of reasoning; see, e.g., Ross and Schroeder (2014, 264-268). Though I think my account of belief can be extended to theoretical reasoning, I shall be neutral on this question. So, I shall speak generally of reasoning, but discuss only practical examples.
[269]   For this kind of case, see Ross and Schroeder 2014, 264-269.

Variation One	Traffic okay.	Traffic bad.
Mad Max	See Mad Max with only a little inconvenience.	See Mad Max with much inconvenience.
Netflix	See next movie in queue.	See next movie in queue.

In Variation Two, by contrast, I factor into my reasoning the possibility that my car would break down on the way. In this variation, my reasoning is more complicated:

Variation Two	Traffic okay; car okay.	Traffic okay; car breaks down.	Traffic bad; car okay.	Traffic bad; car breaks down.
Mad Max	See Mad Max with only a little inconvenience.	Terrible day.	See Mad Max with much inconvenience.	Even more terrible day.
Netflix	See next movie in queue.	See next movie in queue.	See next movie in queue.	See next movie in queue.

We can describe the difference between these two variations by saying that in Variation Two, I do not rely on the proposition that my car would not break down on the way. For I factor into my reasoning the possibility of the negation of that proposition. In Variation One, by contrast, I do rely on the proposition that my car would not break down on the way. In not factoring into my reasoning the possibility of its negation, I rely on the proposition as true in my reasoning, and not merely as highly probable.

For another kind of case sometimes used to illustrate reliance on a proposition as true, consider an example from John Hawthorne.[270] Suppose that I reason as follows:

    (1) My lottery ticket is a loser.
    (2) So, I get nothing if I keep it.
    (3) But I could get a penny if I sell it.
    (4) So, I should sell it.

If I reason in this way, I am relying on (1) as true in my reasoning, and not merely as highly probable.

---

[270]   Hawthorne 2004, 29.

The second constraint puts the third constraint into sharp relief. The second constraint appears to imply that if one believes p, one would be willing to risk anything on p. For, in relying on p as true, one ignores the possibility of p's negation. But now suppose a demon appears and offers me \$1 if p is true against eternal torment if p is false. Ignoring the possibility of p's negation, shouldn't I be willing to take that bet? What have I to lose?

However, it is possible to believe p even if there are some things one would not be willing to risk on p. For example, I believe Los Angeles is in California. But if a demon were to offer me the above kind of bet, it would be very natural for me to reply that though I decidedly believe Los Angeles is in California, I am not so incredibly confident of that proposition for me to feel comfortable taking such a bet.[271] Another way to put the point is that it is possible to believe p even if one does not have a credence in p of 1.[272] And if this is possible, we should expect it to be possible to believe p even if there are some things one would not be willing to risk on p.

Those who emphasize that to believe p is to rely on p's truth are wont to offer a certain kind of clarification. To believe p is to rely on p's truth, as Wedgwood phrases it, "for at least all *normal* practical purposes".[273] Or, as Ross and Schroeder say, "believing that p *defeasibly* disposes the believer to treat p as true in her reasoning".[274] These clarifications are related: a case in which the defeasible disposition is defeated, and so does not operate on default, is precisely a case which is not normal. The idea here is that belief in p involves reliance on p as true in *normal* cases, which allows that there may be some things one would not be willing to risk on p, even

---

[271]   See Wedgwood 2012, 321.
[272]   Frankish 2009, 79.
[273]   Wedgwood 2012, 321.
[274]   Ross and Schroeder 2014, 267-268, italics mine.

though one believes *p*. An account of belief should make good on this idea. So much for the third constraint.

The fourth constraint is that belief is distinct from *acceptance* as that notion has been popularized by L. Jonathan Cohen and Michael Bratman.[275] To accept *p* is to rely on *p* as true in a narrow, specified range of situations.[276] The case of a lawyer is often used to illustrate the notion. The lawyer may know full well that her client is guilty, in light of the evidence. Nevertheless, she may reason as if her client is not guilty in all situations relevant to defending her client in court. In this case, the lawyer accepts her client is not guilty. But she does not believe it. So, belief and acceptance are distinct.

I turn now to the fifth and final constraint. Consider an example. I look around, and I see no elephant in my room. Thus, I have a high credence that it is not the case that there is an elephant in my room. And because of this, it is *difficult*, to say the least, for me to believe there is an elephant in my room. This example illustrates a general connection between belief and credence: a high credence in *p*'s negation makes it difficult to believe *p*.[277] This is the fifth constraint.

Let me summarize the five constraints. First, belief is an enduring state. Second, belief in *p* involves relying on *p* as true. Third, it is possible to believe *p* even though there are some things one would not be willing to risk on *p*. Fourth, belief is distinct from acceptance. Fifth, a high credence in *p*'s negation makes it difficult to believe *p*.

---

[275]   See Cohen 1989 and Bratman 1992.

[276]   This terminology is not universal, though. For example, for Frankish (2004, Chapter 5), to accept *p* is simply to rely on *p* as true in some sense or other. So, in Frankish's terminology, reliance on *p* as true in a narrow range of cases and belief in *p*, which is reliance on *p* as true in normal cases, are both subspecies of acceptance. I thank an anonymous referee for this point.

[277]   One might even think, *a la* Bernard Williams (1973), that it is impossible to believe *p* if one has a high credence in *p*'s negation. But that is more controversial, unless a suitable sense of 'impossible' is specified.

§8.3   Meeting the Constraints

In this section, I develop an account of belief with an eye to meeting each of the five

constraints. At the end of the section, I consider some objections.

§8.3.1 Belief is an Enduring State

To see that an activity can ground an enduring state, recall my

definition of activity types	An event type E is an activity type if and only if there is some way of dividing E into temporal parts $e^1 \ldots e^n$ such that (i) each of $e^1 \ldots e^n$ is a specific way for part of E to occur in an interval of time and (ii) each of $e^1 \ldots e^n$ is a candidate for being an act.

As we saw in §6.4, on this definition, activities such as breathing, which include periods during

which one is dreamlessly asleep, qualify as activities. This definition makes room for the

following kind of view. By virtue of engaging in the activity of climbing to the top, a climber

who is asleep on a cot is in a particular state which disposes her to continue upwards in certain

characteristic ways. This state is the intention to climb to the top, which is an enduring mental

state.[278] And the same story could hold in the case of belief. Perhaps, by virtue of engaging in a

certain kind of activity, a sleeping person is in a state which disposes her to reason in

characteristic ways. And perhaps this enduring state is belief.

§8.3.2 Organizing One's Attention As If P is True

Let us consider what kind of activity might ground belief. Recall the second constraint

from §8.2, namely that belief in *p* involves relying on *p* as true in one's reasoning. This

---

[278]   Philosophers who hold intention is grounded in activity include Ferrero 2002, Thompson 2008, and Moran and Stone 2011. Admittedly, whether mental states like intention and belief are best understood in terms of underlying activities warrants more discussion. For example, Bratman (1987, 4) raises the objection that it seems that one can have an intention to do something before doing anything about it. For a response to this kind of objection, see Ferrero (2002, 173-203).
     As we shall see, my definition of activities makes the view that belief is grounded in an activity a view which can vindicate ordinary talk of choosing beliefs. For my definition bottoms out in actish phenomenal quality, which would explain why ordinary folk talk of choosing beliefs. Not every conception of activity can explain this linguistic data. For example, Boyle (2011, 19-23) develops a conception of activity on which holding a belief is an activity if one holds it because one regards it as rationally correct. But consider the belief that the United States is no longer a colony of Great Britain. Even if the ground by virtue of which I hold this belief is that I regard it as rationally correct, I would have no tendency to say I chose to hold it.

constraint concerns the core role of belief in our mental economy, and so appears to be a natural starting point from which to address this question.

Remember the Netflix or Mad Max case. In this case, I am deliberating about whether to watch Mad Max in theatres or instead stay home and watch Netflix. I have a credence of .6 that traffic would not delay me more than 30 minutes, and a credence of .99 that my car would not break down on the way to the theatre. In both variations of this case, I factor into my reasoning the possibility that traffic would delay me more than 30 minutes. In Variation One, I do not factor into my reasoning the possibility that my car would break down, ignoring it instead. But I do factor this possibility into my reasoning in Variation Two.

Notice the phrases it was natural to use to describe these variations. We say that I *factored a possibility into my reasoning* or instead that I *ignored it*. The opposite of ignoring something is paying attention to it in a particular way. This is what it is to factor something into one's reasoning. In light of this, consider the hypothesis that to rely on *p* as true is to organize one's attention in a certain way. In Variation One of the Netflix or Mad Max case, I attend to, as relevant to my reasoning, only possibilities consistent with my car's not breaking down. So, more specifically, perhaps to rely on *p* as true is *to engage in the activity of attending only to possibilities consistent with p as relevant to one's reasoning*. For brevity, I shall call this activity *organizing one's attention as if p is true*. To illustrate this terminology: in Variation One, I organize my attention as if the proposition that the car would not break down is true.

Let me offer some clarifications. A possibility can occur to one without one factoring it into one's reasoning. For example, in Variation One, the possibility that my car would break down may occur to me, but I do not factor it into my reasoning. Factoring a possibility into reasoning requires not only entertaining it, but also dealing with it as relevant to one's reasoning.

This does not, of course, require believing that it is relevant.[279] This becomes apparent once we say more about what it is to attend to something as relevant to one's reasoning. To do this is to shift one's attention in order to structure what factors into one's reasoning, and what instead either fades away or becomes unheeded background noise. At an appropriate level of generality, attending in this way has a parallel in perceptual experience. If I see a Necker cube, I can shift my attention so that a particular corner pops out to me as in the foreground.[280] Similar comments hold with respect to Duck/Rabbit pictures. In all of these cases, we can shift our attention to organize the manner in which something is before our minds, which structures how we interact with it, whether in perceptual experience or in reasoning. But it is not as if, in shifting attention back and forth when looking at the Duck/Rabbit picture, one must alternate between believing it is a picture of a duck and believing it is a picture of a rabbit. Relevance in attention need not be believed relevance.[281]

Variation One of the Netflix or Mad Max case highlights a particular interval during which I organize my attention as if my car would not break down. But my organizing my attention in this way can extend indefinitely into the future, similarly to how my digesting meals, my breathing, and my securing financial solvency for my family can extend indefinitely into the future. Organizing my attention in this way is an event which can include some temporal parts during which I am asleep.

---

[279]   No account of belief should maintain that, to deal with possibilities in these ways, one must *believe* they are relevant, or not relevant. As dealing with possibilities in these ways is part of belief, to countenance this kind of requirement would set off a regress of beliefs one would need to have, in order to have any particular belief. For example, to believe *p*, one would need to believe (something like) that the possibilities consistent with *p* are relevant and the possibilities not consistent with *p* are not relevant. But to believe *that*, one would have to believe that the possibilities consistent with *that the possibilities consistent with p are relevant and the possibilities not consistent with p are not relevant* are relevant and the possibilities not consistent with *that the possibilities consistent with p are relevant and the possibilities not consistent with p are not relevant* are not relevant. And so on *ad infinitum*. This regress appears vicious, very quickly involving extremely complex beliefs.

[280]   For discussion of Necker cubes, see Jack Lyons (2009, 43).

[281]   On this count, I follow Arvidson 1996 and Watzl 2011. Each develops accounts of attention on which focal objects and contextual objects are presented as relevant to each other in specific ways. And each construes the presentation of relevance as experiential, not doxastic. Consider an example from Watzl (2011):

> In many cases, when you are attending to the saxophone, the sound of the piano is experienced as relevant for or close to the experience of the melody played by the saxophone. (p. 156)

Moreover, organizing one's attention as if p is true is an activity. Recall, an event type is an activity type if and only if there is some way of dividing it into temporal parts such that each of those parts is a specific way for part of that event type to occur in an interval of time, and each of those parts is a candidate for being an act. Organizing one's attention as if p is true is an event type which meets these criteria. For example, there is a sense in which it is possible for me, at any particular time, to consider the possibility that my car would break down, and to experience both paying attention to this possibility and ignoring it as available to me. In this case, whatever alternative I realize, the result is a shift of attention with actish phenomenal quality. And if I ignore the possibility as not relevant, this act is a temporal part of organizing my attention as if my car would not break down. Thus, in general, organizing one's attention as if p is true can be divided into temporal parts each of which is a candidate for being an act.

Per my discussion of activities in §6.4, the relevant sense in which it is possible for me to shift my attention as described concerns only the event type, not the event token with all its details. Thus, in the sense of possibility which is relative to particular individuals at particular times, it may not be possible for one to organize one's attention as if p is true at a given time. For example, one's evidence against p may be overwhelmingly strong. I shall discuss this issue in more detail in §8.3.5, where I explain why a high credence in p's negation makes it difficult to organize one's attention as if p is true.

Attention is commonly divided into the voluntary and the involuntary.[282] The person controls voluntary attention, as when one focuses on one of multiple visual stimuli to which one could pay attention. Or, I may voluntarily shift my attention when looking at a Necker cube or a Duck/Rabbit picture. But factors other than the person control involuntary attention. For

---

[282] Voluntary attention is also called endogenous or top-down; involuntary attention is also called exogenous or bottom-up. For this distinction, see Ruff (2011, 5), Mole (2011, 67), Armstrong (2011, 87-88), and Wu (2011, 107).

example, the nearby lightning strike may grab one's attention by force. This way of dividing attention fits nicely with the view that a way of organizing one's attention is an activity, in my sense. On the one hand, when a temporal part of the organization of one's attention is controlled by factors other than the person, the sense in which it is not possible for this temporal part to be an act is the sense relative to particular persons and times. On the other hand, when the person controls a temporal part of such organization, this part is, plausibly, an act. Moreover, so long as the parts are suitably bite-sized, there is no temporal part of such organization which we can rule out from the possibility of being an act simply on account of the event type it is. That is, these parts are candidates for being acts. And this is to say that ways of organizing one's attention are activities.

We began by pointing out that, in our mental economy, the core role of belief in p is reliance on p as true. Indeed, this is the second constraint from §8.2. I have suggested that to rely on p as true is to organize one's attention as if p is true. We have seen that organizing one's attention in this way is an activity. With all of this on the table, the core proposal of this chapter is that belief in p is the state one is in by virtue of engaging in the activity of organizing one's attention as if p is true.

§8.3.3 The Qualified Nature of Activities

Recall the third constraint on an account of belief from §8.2, namely that it is possible to believe p even if there are some things one would not be willing to risk on p. For example, I believe Los Angeles is in California. But if a demon were to offer me $1 if this proposition is true against eternal torment if it is false, it is consistent with my believing it that I turn this bet down. Now, if belief in p is grounded in the activity of organizing one's attention as if p is true, it is not immediately clear how this constraint can be met. For one might think that anyone

engaged in this activity would ignore the possibility that Los Angeles is not in California, and so would be willing to take the demon's bet.

A solution to this problem derives from a general fact about activities to which we have yet to give attention. This is the fact that activities are *qualified* in nature.

Take the example of the person climbing to the top of the mountain. If asked what she is doing, she would say that she is climbing to the top. I shall take this kind of talk at face value. So, I assume she is indeed engaged in the activity of climbing to the top, even if she does not complete this activity. Now suppose that she comes to find, to her surprise, that the metal rings to which she had expected to secure her ropes are no longer there. Upon seeing the absence of the rings, she turns around in frustration.

Compare this case to one where she finds the metal rings as expected, but turns around because she starts to feel it would be nice to go river rafting before the day is over. It is more natural to say that she changes her mind about whether to complete the climb in this second case than it is to say this in the first. Moreover, it is more natural to say that she is prevented from completing the climb in the first case than it is to say this in the second. This means that in the first case, in which the metal rings are gone, there is a sense in which she does not change her mind about whether to complete her climb, but instead is prevented from completing it. This is the case even if she could, in principle, make the extremely dangerous attempt of climbing the rest of the way without ropes. This observation indicates that she is *not* engaged in what we might call the activity of climbing to the top *no matter what*. Were she engaged in this other activity, she would have to change her mind in order to turn back at the absence of the metal rings. Climbing to the top no matter what is an *unqualified* activity, and much more desperate than what we would normally call climbing to the top. As it is, she is engaged in a *qualified* activity. And hers is the normal case. Generally, activities are qualified.

As we might put it from a third-person perspective, our hero is climbing to the top *unless certain conditions obtain*. These are conditions to which she can respond, when face-to-face with them. As Luca Ferrero might put it, these are conditions that make finishing the climb either no longer feasible or no longer advisable.[283] But not to be misunderstood, it is not as if the climber can complete what she is doing either by getting to the top, or by paying someone to remove the metal rings before she gets to the relevant point.[284] Instead, she pursues the whole activity "against the backdrop" of conditions on which climbing to the top is feasible and advisable: if this backdrop is pulled out from underneath her, she is prevented from completing the activity.[285]

Climbing to the top is an activity with a built-in end point. But activities which extend indefinitely into the future are also typically qualified. Consider, for example, the activity of securing financial solvency. This activity involves promoting financial solvency in various situations. But it is not the activity of promoting solvency *no matter what*. There may be situations in which promoting solvency would be infeasible or inadvisable. Thus, securing financial solvency is qualified in the sense that it is the activity of promoting solvency in a certain *range*, but not *all*, of the situations in which it is possible to do so.

The same is true of organizing one's attention as if p is true. This is not the activity of organizing one's attention as if p is true *no matter what*. Similar to securing financial solvency, it is the activity of attending only to possibilities consistent with p in a certain range, but not all, of the cases in which it is possible to do so. Abnormal cases, including that of the demon's bet, fall outside this range. Thus, to organize one's attention as if p is true is to attend only to possibilities consistent with p in all *normal* cases. At least, that is how we might put it from a third-person

---

[283]   Ferrero 2009, 722-725.
[284]   On the approach I am exploring, the activity in this example grounds one's having an intention. In this connection, the content of the intention is not a material conditional. The intention is qualified in a way, but its content is *not* the following: either get to the top or make it infeasible or inadvisable to do so. On this point, see Ibid., 703-709.
[285]   Ibid., 705.

perspective. This distinction between normal and abnormal cases reflects the qualified nature of the activity.

So, the view that belief is grounded in organizing one's attention as if p is true can explain why it is possible to believe p even if there are some things one would not be willing to risk on p. In the abnormal case in which the demon offers an insane bet on p, one can attend to the possibility of p's negation as relevant to one's reasoning and still count as organizing one's attention as if p is true. For this activity, like activities generally, is qualified.

§8.3.4 The Range of Normal Cases

Recall the fourth constraint on an account of belief from §8.2, namely that belief is distinct from acceptance as that notion has been popularized by L. Jonathan Cohen and Michael Bratman.[286] Consider, again, the example of the lawyer. She may know full well that her client is guilty. But she may reason as if her client is not guilty in all situations pertinent to defending her in court. In doing so, she accepts that her client is not guilty, but she does not believe it.

If my account of belief is to meet this constraint, it must be that the lawyer does not organize her attention as if her client is not guilty. What we said in §8.3.3 is of help on this front. True, the lawyer does attend only to possibilities consistent with her client's not being guilty in all cases pertinent to defending her in court. But she does not attend only to these possibilities in all normal cases. And to organize one's attention as if p is true is to attend to only to these possibilities in all normal cases.

The range of normal cases is more expansive than the cases within which one who accepts p attends only to possibilities consistent with p. To accept p, one must preconceive a

---

[286] See Cohen 1989 and Bratman 1992.

precise boundary within which one will attend only to possibilities consistent with *p*. The case of the lawyer is a prime example of this fact.

By contrast, it is entirely possible to organize one's attention as if *p* is true without preconceiving any such boundary. This is a feature of activities in general, including qualified activities. For example, we can suppose that the climber does not preconceive of the precise conditions which would cause her to turn back, though she can recognize them when she sees them. This is why we were careful to say that it is a third-person perspective, not necessarily her own, from which we might describe her activity as *climbing to the top unless certain conditions obtain*. She sallies forth into the world with boldness, in engaging in her activity. This same boldness is possible in organizing one's attention as if *p* is true, though one can recognize abnormal cases like the demon's bet when one sees them. Due to this boldness, among the cases in which one will attend only to possibilities consistent with *p* are cases in which it is highly desirable to take truths, but *only* truths, as premises in one's reasoning.[287] In these cases, one who organizes her attention as if *p* is true will attend only to possibilities consistent with *p*, but the lawyer would not similarly attend only to possibilities consistent with her client's not being guilty. So, on my account, belief is distinct from acceptance.[288]

§8.3.5 Why Belief in P is Difficult with a High Credence in P's Negation

Recall the fifth and final constraint from §8.2: that a high credence in *p*'s negation makes it difficult to believe *p*. For example, I have a high credence that it is not the case that there is an elephant in my room. And because of this, it is difficult for me to believe there is an elephant in my room.

---

[287] Compare Frankish 2009, 87.
[288] Similarly, on my account, belief is distinct from supposition. When supposing *p* (as in reasoning by *reductio* to *p*'s negation), one preconceives a precise, narrow boundary within which to attend only to possibilities consistent with *p*. Normal cases outstrip this boundary.

Before explaining how my account meets this constraint, recall my account of credence from §7.4.1. Credence in p is a feeling of p being true, which has a particular degree of strength, which may or may not be conscious at the time, and which helps explain certain dispositions, including the utility-risking dispositions one has concerning p given one's preferences. It is helpful to highlight two of my assumptions about credence, assumptions which dovetail nicely with this account. My first assumption is that among the dispositions explained by credences are dispositions to attend to p-possibilities[289] as relevant to one's reasoning if one has a positive credence in p. My second assumption is that these dispositions do not necessitate engagement in the activity, even if one is in the relevant circumstance. This second assumption is important in light of my account of belief. If a positive credence in p's negation *necessitated* attending to not-p-possibilities as relevant, it would not be possible to organize one's attention as if p is true if one has a positive credence in p's negation, however small. But it is entirely possible to believe p despite having a positive credence in p's negation.

On my assumptions, a positive credence in p's negation disposes one to attend to not-p-possibilities as relevant to one's reasoning. This disposition does not necessitate. However, plausibly, the higher one's credence in p's negation, the *stronger* a disposition it is to attend to not-p-possibilities as relevant. Moreover, it is plausible that at some point, the disposition is so strong that in the sense of possibility relative to particular persons and times, it is not possible for one to ignore these possibilities in one's reasoning.

To see why this should be plausible, remember that a credence, being a feeling, is not a state that is grounded in an activity. But credence does dispose us to engage in particular activities. And we can compare credence to other feelings. For example, we all have strong

---

[289]   A p-possibility is a possibility in which p is true, and a not-p-possibility is a possibility in which p's negation is true.

feelings of self-preservation, which manifest vigorously when we are extremely hungry or near the edge of a cliff. These feelings make a long-term fast difficult, but eating daily easy. More to the point, they make it so that, in the sense of possibility relative to particular persons and times, it is not possible for me to fling myself off a cliff. Nor is it possible, in this sense, for me to *avoid* engaging in some activities, such as the activity of keeping myself away from fast moving traffic.

We can generalize. States not grounded in activities, if sufficiently strong, can make it impossible to engage in certain activities, in the sense of possibility relative to particular persons and times.[290] When not thinking clearly about such states, we may think that, if given sufficient incentive, we could just do anything we can coherently imagine. But that is fantasy: such states make certain activities genuinely available to us and effectively close off others.[291]

On my account of belief, to believe *p* is to organize one's attention as if *p* is true. And to do that requires ignoring not-*p*-possibilities in one's reasoning. But as we have seen, it is plausible that a high credence in *p*'s negation disposes one so strongly as to close off the possibility of ignoring not-*p*-possibilities, in the sense of possibility relative to particular persons and times. And this is why, on my account, a high credence in *p*'s negation makes it difficult to believe *p*. So, my account meets the fifth and final constraint from §8.2.[292]

---

[290]   As in §6.4, this sense of possibility is to be distinguished from that which concerns simply the *event type*. In this other sense of possibility, the event type *flinging myself off a cliff* can be an act; that is, it is a candidate for being an act.

[291]   Compare Mourad 2008, 61.

[292]   I believe this explanation of the fifth constraint has attractions in comparison to others on the market. Frankish (2004, 150) and Wedgwood (2012, 324-328) hold that it is difficult to believe *p* if one has a high credence in *p*'s negation because it is *irrational* to do so. But this gives us an explanation only if it is always difficult to do what is irrational. And this is not clear. Sometimes doing what is rational is more difficult than doing what is irrational. And even if irrationality always presents a kind of difficulty, I appear quite able to irrationally eat too much ice cream in a sense in which I am not able to, say, believe there is an elephant in my living room.

Another possible explanation appeals to credal reductivism. Credal reductivists could maintain that no belief in *p* can reduce to an overall credal state which includes a high credence in *p*'s negation. Then they could explain why it is difficult to believe *p* if one has a high credence in *p*'s negation by saying that this combination of states is impossible. But as Ross and Schroeder (2014, p. 268-271) argue, credal reductivism faces a dilemma. Either belief in *p* reduces to a credal state which includes a credence of 1 in *p*, or it reduces to a credal state which includes a credence of less than 1 in *p*. If the former, we can explain why belief in *p* involves relying on *p* as true, and not merely as highly probable. But then we cannot explain why it is possible to believe *p* even though there are some things one would not be willing to risk on *p*. For one with a credence of 1 in *p* would be willing to risk anything on *p*. On the other hand, if belief reduces to a credal state which includes a credence of less than 1 in *p*, then we cannot explain why belief in *p* involves relying on *p* as true, and not merely as highly probable. For to have a credence in *p* of less than 1 is not necessarily to rely on *p* as true.

§8.3.6 Objections and Replies

In this section, I reply to three objections to my account.

First objection. It is possible to believe *p* without ever having entertained *p*. Consider, for example, the proposition that one cannot fly unaided. One might believe this proposition, but never consider the matter. But then whatever believing this proposition amounts to, it does not seem to be grounded in an activity of organizing one's attention in a particular way.[293]

In response, my definition of activities allows for the possibility that one engages in the activity of organizing one's attention as if *p* is true even if one has never entertained *p*. On my definition, an event is an activity just in case there is some way of dividing it into temporal parts such that each of these parts is a specific way for part of the event to occur, and each is a candidate for being an act. There are many specific ways for part of the event of organizing one's attention as if *p* is true to occur in which one does not entertain *p*. One could be reasoning about something unrelated to *p*, one could be asleep, and so on. Moreover, one who has never entertained *p* may nevertheless have the ability to, at any given time, entertain *p*- and not-*p*-possibilities and act to attend to them in the relevant way. Thus, an event of organizing one's attention as if *p* is true could meet the conditions my definition lays down, even if one has never entertained *p*.

One might object that if one has never entertained *p*, one would not have the ability to, at any given time, entertain *p*- and not-*p*-possibilities. For our purposes, this may well be.[294] But if

---

[293]   I thank an anonymous referee for raising this objection.

[294]   By my lights, it is not obvious that it is possible to believe *p* without ever having entertained *p*. I hold, with Robert Audi (1994), that we do well to distinguish believing *p* from the disposition to believe *p*. And it might be that we are prone to mistake cases in which one has a disposition to believe *p* without ever having entertained *p* for cases in which one believes *p* without ever having entertained *p*. In any case, my account leaves as an open question whether one can believe *p* without ever having entertained *p*. The answer to the question depends on what it takes to have the ability to, at any given time, entertain *p*-possibilities, not-*p*-possibilities, and act to attend only to *p*-possibilities as relevant to one's reasoning. Plausibly, this ability requires possession of each concept needed for entertaining *p*. This, in itself, is a nice result, for it allows my account to explain why no proposition incomprehensible to one is a proposition one believes. In addition, the relevant ability may be more sophisticated than the raw capacities which allow us to entertain *p* for the first time, and may require *p*'s having been entertained in the past by means of these other capacities. For comparison, when we say that one has the ability to selectively isolate one's ear muscle, so as to wiggle it, the ability we refer to seems to be more sophisticated than the raw capacities which allowed that person to isolate and wiggle her ear for the very first time.

one does not have this ability, I submit that she also fails to have belief in *p*. Thus, to place

having entertained *p* as a necessary condition on this ability is to jeopardize the first objection's

premise that it is possible to believe *p* without having entertained *p*.

Perhaps most importantly, one might worry that my definition categorizes too many

events as activities, in allowing for organizing one's attention as if *p* is true to be an activity even

in cases in which one has never entertained *p*.[295] Now, while an activity is a doing in some sense,

it need not be a doing in every sense. For example, there appears to be a sense in which someone

who is asleep is doing nothing at all. But as mentioned in §6.4, when focusing on complex events

of significant duration and taking language at face value, there is a sense in which, even when

asleep, one might be climbing to the top of the mountain, traveling up the west coast, and getting

a PhD. Moreover, there is a sense in which these events are not simply happening to one. One is

doing them. With this in mind, rather than showing my definition to be too permissive, we can

learn a different lesson from the first objection. This is the lesson that, in the sense of 'doing' at

issue, one can be doing something without *ever* being aware of it. I recently saw a physical

therapist who informed me that I walk with my right foot turned further outward than my left

foot. I might never have become aware of this. But even so, my walking in this way was

something I was doing, not something happening to me. My definition categorizes this as an

activity, and, as with organizing one's attention as if *p* is true where one has never entertained *p*,

it is right to do so.

Second Objection. Suppose I am reasoning about what to do today. I think to myself that

cycling in the park would be easy and inexpensive. But my enthusiasm wanes after I remember a

handful of things I believe. I first remember that my bike has a flat. I wonder if I could use my

---

[295]   I thank an anonymous referee for this point.

kit for fixing tires, but then I remember that my friend borrowed it. I consider calling her, but then I remember that she is unavailable, being out of town. So, the possibility that cycling in the park would be easy and inexpensive is inconsistent with things I believe. Nevertheless, I paid attention to this possibility in my reasoning. So, contrary to my account, it is not the case that, to believe something, I must attend only to possibilities consistent with it as relevant to my reasoning.[296]

In response, the possibility that cycling in the park would be easy and inexpensive is consistent with each of the handful of propositions, taken individually. For example, it is consistent with my bike having a flat, as I am handy with my kit for fixing tires. What the possibility at issue is inconsistent with is this conjunction: my bike has a flat and my friend borrowed my kit for fixing tires and my friend is unavailable. Now, one can believe each of a set of propositions without believing their conjunction. So, the example makes trouble for my account only if, when I am paying attention to the possibility that cycling in the park would be easy and inexpensive, I already believe the relevant conjunction. But, plausibly, the example is precisely a case in which I do not believe this conjunction, and instead merely its conjuncts. It looks to be a case in which, put colloquially, I am putting things together. That is, it looks to be a case in which I reason from a set of things I believe individually to a conjunction of those things. And it is precisely when I arrive at belief in that conjunction that I cease to pay attention, in my reasoning, to the possibility that cycling in the park would be easy and inexpensive. Rather than being contrary to my account, this is precisely what my account predicts.

Third objection. Suppose a healthy friend of mine has a practice of paying attention, each day, to the possibility that she will have a heart attack that day. She does this to motivate herself

---

[296]   I thank an anonymous referee for raising this objection.

to continue going to the gym. Now even so, it might well be that, each day, she believes she will not have a heart attack that day. And that is contrary to my account.[297]

In response, while my friend does indeed attend, each day, to the possibility that she will have a heart attack that day, she does not attend to that possibility *as relevant to her reasoning*. Instead, she attends to that possibility to motivate herself, that is, to instill or invigorate certain feelings. Similarly, I may imagine a possibility in which I do not have shelter, to increase my gratitude for my apartment. In neither case is the possibility attended to as relevant to one's reasoning. Otherwise, I would start considering what could ensure I have shelter by tonight, and my friend would start considering whether to stop her normal workout routine so as to make the most of what could be her last hours. Attending to a possibility as relevant to one's reasoning and attending to a possibility in order to affect one's feelings both involve attending to that possibility, but in different ways.

In any case, not only does my account meet each constraint from §8.2, it also stands up to objections.

§8.4   Benefits of My Account

In this section, I argue that my account of belief can explain features of belief it is unclear rival accounts can explain. In particular, my account can resolve a tension in our conception of belief, and in the process vindicate ordinary talk and blaming practices that suggest we often choose our beliefs. In addition, my account can explain, in a simple and elegant way, why it can be perfectly coherent to believe *p* without having maximal credence in *p*.

---

[297]   I thank an anonymous referee for raising this objection.

§8.4.1 Resolving the Tension in Our Conception of Belief

There appears to be a tension in our conception of belief. On the one hand, belief seems

to be at the mercy of evidence. As William Alston expresses the idea:

> Can you, at this moment, start to believe that the United States is still a colony of
> Great Britain, just by deciding to do so?....It seems clear to me that I have no such
> power.[298]

Of course, often our evidence is less conclusive. But even in those cases, it does not appear we

can alter our degree of belief by choice.[299]

On the other hand, we talk as if we often choose our beliefs. As Alex Worsnip has

pointed out, the phrase 'I choose to believe' returns over 393,000 separate results on Google.[300]

Worsnip suggests this talk is especially common in the cases of religious belief, belief in the

testimony of a friend, belief in one's own abilities, and the termination of deliberation in belief

instead of suspension.[301] In my own case, I need only recall the last family reunion for

confirmation that we do talk this way.[302] Moreover, it does not appear that this talk is shorthand

for one's choosing to do something distinct from believing which, in some manner or other,

causes belief.[303] And notice that, in lockstep with ordinary talk of choosing to believe, is the

intuition that a belief is the kind of thing for which one can be appropriately blamed. Consider,

for example, a college student who believes that people with dark skin color are inferior.[304]

---

[298]   Alston 1989, 122.
[299]   Ibid., 123-127.
[300]   Worsnip 2015, 233.
[301]   Ibid., 233-236.
[302]   An anonymous referee has suggested that ordinary folk who talk of choosing to believe are mistaking *acceptance* for belief. To accept *p* is to rely on *p* as true in a narrow, specified range of situations, as when a lawyer, aware that her client is guilty, reasons as if her client is not guilty in situations related to defending her client in court. It is generally agreed that one can choose to accept *p*. However, in many instances of the cases Worsnip mentions, it is unlikely ordinary folk are mistaking acceptance for belief. For, in many instances of these cases, it is not merely a narrow, specified range of situations in which the person relies on *p* as true. For example, most people who say they chose to believe that God exists tend to rely on that proposition as true in general or normally, not just in a narrow, specified range of situations. Notice the contrast between their case and the case of the lawyer.
[303]   Generally, those who talk of choosing to believe do not take themselves to refer to choosing to do something else, which causes belief. Notice that, in general, it would be infelicitous to ask the speaker what she chose to do, distinct from believing, which caused her belief. She may talk of choosing to believe for certain reasons, but have nothing to say if pressed to identify something distinct from believing which she chose to do to cause her belief. Contrast this with talk of choosing to have a tattoo. It is fine to ask the speaker what she chose to do, distinct from having the tattoo, which caused her to have the tattoo. She might speak of choosing to go to a tattoo parlor, or etc.
[304]   For this example, see Nottelmann 2007, 3.

Intuitively, this student is blameworthy for holding this racist belief. But if belief is not the kind of thing one can ever choose, it is puzzling how the student could be blameworthy for her belief.[305]

So, it seems belief is at the mercy of evidence, yet our ordinary talk and blaming practices suggest we often choose our beliefs. How should we resolve this tension? Is belief always at the mercy of evidence, or can we sometimes choose our beliefs?

Now, when philosophers talk of choice, they tend to refer to the forming of an intention or plan. But we need not assume ordinary folk have something this precise in mind. To explain why ordinary folk would say they chose to do something, all we need is the observation that 'choosing' is an ordinary word for selecting between alternatives. Thus, if something's being a selection between alternatives can be part of *what it is like* for one to do it, we should expect folk, in such cases, to report that they chose to do what they did. And this is exactly what holds for exercises of will. An exercise of will is an event wherein one experiences oneself as the source of one among multiple available alternatives. Put differently, this is to experience oneself as selecting between alternatives. We should not be surprised if ordinary folk report an exercise of will as something they chose to do.

With this in mind, we can resolve the tension in our conception of belief. On my account, belief in p is grounded in the activity of organizing one's attention as if p is true. To say that organizing one's attention as if p is true is an activity is to say that its parts are candidates for being exercises of will. So, at any given time, it is possible to experience the availability of the alternative of attending only to possibilities consistent with p, the alternative of doing otherwise, and to realize one of these alternatives. On my account, realizing the first alternative grounds

---

[305]   I take the idea that blameworthiness requires choice, in at least some sense, to be pretheoretical and common. Of course, some philosophers reject this idea. See, for example, Sher 2009 and Vargas 2005.

one's belief in *p* at that time, and realizing the second grounds one's suspension with respect to *p* at that time. So, we should expect ordinary folk who realize the first alternative to be inclined to say that they chose to believe *p* at that time. For at that time, an exercise of will constituted their engagement in the activity which grounds belief in *p*. And we should expect ordinary folk to report exercises of will as things they chose to do.[306]

On the other hand, we can also explain why it seems that belief is at the mercy of evidence. For 'belief' often refers to credence. And credence is at the mercy of evidence, in the relevant sense. There is some manner in which our evidence affects which credence in *p* we have. And however our evidence does this, it is not by motivating an activity. This is because, on my account, credence is not grounded in an activity. Thus, however evidence impacts credence, it bypasses our activities in doing so. In this sense, credence is at the mercy of evidence.

§8.4.2 Coherent Uncertain Belief

As we saw in §8.2, one constraint on an account of belief is that belief in *p* involves reliance on *p* as true. Another constraint is that it is possible to believe *p* even if there are some things one would not be willing to risk on *p*. As I explained in §8.3, my account of belief meets these constraints.

But other accounts do so as well. For example, according to Ross and Schroeder, to believe *p* is to have a defeasible disposition to evaluate alternatives in the manner by which one would do so conditional on *p*.[307] The manner by which one would evaluate alternatives conditional on *p* is, roughly, the manner by which one would do so if one had a credence of 1 in

---

[306]   Why might it be, as Worsnip has suggested, that talk of choosing to believe is especially common in the cases of religious belief, belief in the testimony of a friend, belief in one's own abilities, and the termination of deliberation in belief instead of suspension? Well, it is typical of such cases that neither one's credence in *p* nor one's credence in *p*'s negation is overwhelmingly high. So, in these cases, one's credal dispositions are typically not overwhelmingly strong. In light of this, we should expect many instances in which both organizing one's attention as if *p* is true and doing otherwise are available to particular individuals at particular times. And, I might add, this is how these cases often seem like from the inside.

[307]   Ross and Schroeder 2014, 264.

p and distributed one's probabilities accordingly. When this disposition is in default operation, one relies on *p* as true. However, the default can be overridden, as in the case in which a demon appears and offers $1 if *p* is true against eternal torment if *p* is false. So, having this defeasible disposition is consistent with there being some things one would not be willing to risk on *p*.

Ralph Wedgwood has raised a problem that, *prima facie*, applies to Ross and Schroeder's account. On their account, one who believes *p* in a normal situation evaluates alternatives in the way one would do so conditional on *p*. But to evaluate alternatives in this way is to treat *p* as though it were *certain*. At the same time, their account allows that one who believes *p* in a normal situation can have a credence of less than 1 in *p*. But to have a credence of less than 1 in *p* is to treat *p* as though it were *not certain*. Thus, so the objection goes, anyone in a normal situation who believes *p* with a credence of less than 1 in *p* is *incoherent*. As Wedgwood puts it, on this kind of account,

> …the threat of incoherence is obvious: by having an outright belief in a proposition *p* of which one is not maximally confident, one is in a way simultaneously treating *p* as though it were certain and also treating *p* as though it were not certain.[308]

But it is absurd to think that anyone in a normal situation who believes *p* with a credence of less than 1 in *p* is thereby incoherent.

Wedgwood offers a way to avoid the incoherence objection: countenance two different kinds of credences, namely *theoretical* credences and *practical* credences.[309] Roughly, the functional role of theoretical credence is to track evidence, while the functional role of practical credence is to guide our deliberation about what to do. Wedgwood proposes that to believe *p* is to be disposed to have a practical credence of 1 in *p* in normal situations. Since the disposition is to have a practical credence of 1, it qualifies as a way of relying on *p* as true. And since the

---

308   Wedgwood 2012, 318.
309   Wedgwood 2012.

disposition is restricted to normal situations, one who believes p need not be willing to take the demon's bet. At the same time, it is clear that (theoretically) uncertain believers in normal situations are entirely coherent. For it is entirely coherent to have a practical credence of 1 in p and yet a theoretical credence of less than 1 in p.

Wedgwood's account explains why uncertain believers are coherent. But it commits us to the view that we have two very different credal systems. And it would be nice if we could explain why uncertain believers are coherent without taking on this commitment.

Here my account can help. On my account, belief in p is the state one is in by virtue of engaging in the *activity* of organizing one's attention as if p is true. This qualifies as "treating p as though it were certain" only if this phrase refers to an activity. But on my account, credence is *not* a state one is in by virtue of an activity. Recall, again, my characterization of credence: A credence in p is a feeling of p being true, which has a particular degree of strength, which may or may not be conscious at the time, and which helps explain certain dispositions, including the disposition to attend to p-possibilities as relevant to one's reasoning if one's credence in p is positive. And remember, from §7.5.1, that feelings are essentially passive: it is not possible to experience oneself as the source of having a definite feeling of a particular strength in p's truth, in the midst of other alternatives experienced as available to one. So, it is not the case that a credence is grounded in an activity. In any case, it follows that one's having a credence of less than 1 in p qualifies as "treating p as though it were not certain" only if this phrase refers not to an activity, but rather to a disposition not grounded in activity. So, on my account, the uncertain believer engages in the activity of treating p as though it were certain,[310] but treats p as though it were not certain in a different sense.

---

[310]   Recall: on my account, it is a *qualified* activity of treating p as certain; so, one need not be willing to take the demon's bet.

There is nothing incoherent about an agent with these features. Compare with the case of intention. One who intends to keep a diet engages in an activity of treating the cake as not to be eaten. But she may nevertheless treat it as *to* be eaten in another sense, disconnected from activity. For example, perhaps her hankering for the cake is a mental state that, in some sense, treats the cake as to be eaten. But this is not incoherence. It is the human condition.

Similar points hold with respect to uncertain belief. In whatever way a credence of less than 1 "treats" p as though it were not certain, it is a way of treating not grounded in activity just as a hankering for cake is not grounded in activity. And while a credence in p is a feeling which disposes one to attend to p-possibilities, this disposition does not necessitate doing so. Thus, one can ignore these scenarios in some cases, though perhaps only if one's credence is not too high. This is relevantly similar to how a hankering for cake does not necessitate the eating of cake, despite disposing one towards it. The main point is that, in both cases, the subject is not incoherent, but instead simply human.

The human condition described above is not a condition in which one's mind is "in conflict". The dieter may be glad to have a mild hankering for cake, as she may want to avail herself of one in the future, when the situation is appropriate. It is good to suffer feelings, even if we have to actively rule over them. Similarly, it is good for the believer in p to have a positive credence in p's negation. We need credences capable of tracking evidence which, if significant enough, can disrupt our believing. It is good to suffer credal dispositions, even if, at times, we have to actively rule over them. Just as dieting has its place, so does organizing one's attention as if p is true. This activity massively simplifies computational demands, for one, and even appears

indispensable on that front.[311] On this picture, uncertain belief is a perfectly coherent, human state of mind.

So, by applying the distinction between activities and non-activities, my account explains why uncertain believers can be coherent. And we need not posit two very different credal systems, as on Wedgwood's account. Now, I shall not argue that it is only my account which can handle the incoherence objection without multiplying credal systems. To recall, on Ross and Schroeder's account, the uncertain believer in a normal situation evaluates alternatives in the manner by which one would do so conditional on *p*. It might well be possible to explain why the sense of 'treating' in which doing so qualifies as treating *p* as though it were certain is *not* the sense of 'treating' in which having a credence lower than 1 in *p* qualifies as treating *p* as though it were not certain. It is worth mentioning, though, that if this explanation appeals to the activity/non-activity distinction, it will look strikingly similar to my own.[312]

§8.5   Conclusion

In Chapter 7, we concluded that to escape the At the Mercy of Evidence Objection, we need to reject the idea that all epistemic reasons are evidential. As we shall see, what makes a reason epistemic is that it is the "right kind" of reason for belief. It is plausible to think that what makes a reason the right kind of reason for belief depends on the nature of belief. My account of belief makes good on this. On my account, belief is a state grounded in an activity. And it might be that practical reasons and activities are, as it were, made for each other. If so, we should expect that some practical reasons are the right kind of reasons for belief, on my account. That is,

---

[311]   For this point, see Foley (2009, 45) and Ross and Schroeder (2014, 265-267).

[312]   My account of belief also applies to the debate about *doxastic voluntarism*, the view that one can, in some sense, believe at will (for discussion of doxastic voluntarism, see, e.g., O'Shaughnessy 1980, Alston 1989, and Frankish 2007). If my account is correct, a version of voluntarism is true about outright belief, but no version is true of credence. In addition, this version of voluntarism about outright belief is moderate, as one cannot choose to believe *p* if one has a high credence in *p*'s negation. So, my account vindicates a restricted, moderate version of voluntarism which is similar to what Keith Frankish (2007, 527) calls "weak voluntarism".

we should expect that some practical reasons are epistemic. I address this issue head-on in the

next chapter.

Chapter 9: Against Evidentialism

§9.1   Introduction

As I explained at the end of Chapter 6, the only remaining response to the At the Mercy of Evidence Objection is to deny the premise that if an obligation is epistemic, it is an obligation to respond to one's *evidence* in a certain way. As a matter of definition, we can assume that epistemic obligations are obligations to respond to one's *epistemic* reasons in a certain way. So, we can rephrase the premise at issue as the premise that all epistemic reasons are evidential. Let us call the thesis that all epistemic reasons are evidential *evidentialism*. In this chapter, I argue that evidentialism is false.

The chapter is structured as follows. In §9.2, I follow Mark Schroeder and lay out several earmarks or distinguishing features of epistemic reasons as against non-epistemic reasons. In §9.3, with these features in mind, I examine two main arguments in the literature for evidentialism, from Thomas Kelly and Nishi Shah. As I explain in §9.4, to come to terms with these arguments, we shall have to take a stand on what the norm for correctness on belief is. While many hold that a belief is correct just in case it is true, I argue that this is not the full story. Instead, I argue that while a belief is correct if true and incorrect if false, correctness and incorrectness come in degrees, and practical matters accentuate how much correctness or incorrectness a belief has. As I explain in §9.5, taking this stand on the norm for correctness on belief helps us to better understand and appreciate cases given by Schroeder in which, to all appearances, practical reasons are epistemic reasons. This shall complete my argument against evidentialism. I conclude in §9.6.

§9.2   Earmarks of Epistemic Reasons

Let us begin with an example. Suppose you are wondering whether it will rain this weekend. Your local weatherman predicts that it will not. But your friend with whom you would like to hike this weekend offers you $100 to believe it will not rain, perhaps knowing that you will agree to hike only if you think it will not rain.

Epistemologists are wont to say that the weatherman's prediction is an epistemic reason to believe, while your friend's bribe is a non-epistemic, merely practical reason to believe (if a reason to believe at all).[313] Though this thought deserves more unpacking than I will be able to give it, it is natural to think that there is a unique way or manner in which it can be rational for one to believe, which contrasts with ways in which it can be rational to do other things. And an attractive idea is that the weatherman's prediction bears on its being rational, in this way, for you to believe it will not rain, while the same cannot be said of your friend's bribe. The bribe may help make belief rational in some other sense, a sense that is not unique to belief. Suppose that, likely counter to fact, you are able to believe it will not rain on the basis of the bribe. Then the sense in which you would be rational in doing so is exactly the sense in which you would be rational in doing *anything else* to procure $100: tie your shoes improperly, eat only ice cream for lunch, etc. Clearly, this is not a kind of rationality unique to, appropriate to, or suited to beliefs in particular.

So, we recognize an intuitive distinction between epistemic and non-epistemic reasons for belief, and we can identify paradigm examples of each. Evidence is a paradigm example of an epistemic reason, while a bribe is a paradigm example of a non-epistemic reason. And we

---

[313]   Some, like Hieronymi 2006, hold that it is a reason, not for forming a belief, but for *intending* to form the belief.

have this nascent, inchoate idea that there is a kind of rationality – *epistemic* rationality – that only epistemic reasons affect.

But, as it turns out, this kind of distinction turns up in other domains. Suppose that you are offered a bribe to form an intention to consume a particular substance tomorrow, and told that the money will be deposited this evening if you succeed in forming this intention.[314] This bribe feels like it affects a kind of rationality that is not suited to intention in particular, while the fact that the substance is one's favorite meal feels like it affects a kind of rationality that is suited to intention. Or consider the social benefits of admiring a nasty person who happens to have substantial power over you: such admiration may be rational, but not rational, as we might say, *qua* admiration.

Hence, it is natural to think that the distinction between epistemic and non-epistemic reasons is a special case of a more general distinction. In the literature, this general distinction is called the distinction between the 'right kind' and the 'wrong kind' of reasons.[315] Different accounts of it are on offer. One influential account, from Derek Parfit and Christian Pillar,[316] has it that the right kind of reasons are *object-given* while the wrong kind are *state-given*. Object-given reasons, in some sense, bear on the object or content of the attitude, while state-given reasons bear on the attitude itself. The account is taken to imply that evidence is a right kind of reason for belief because it bears on the content of a belief and that a bribe is a wrong kind of reason because it, instead, bears on the desirability of having the belief. Similar points go for intention, admiration, etc. Pamela Hieronymi offers a different account of the distinction, on which a consideration is the right kind of reason for an attitude just in case it bears (or is taken to

---

[314]   Consider, e.g., Kavka 1983.
[315]   See Wlodek Rabinowicz and Toni Ronnow-Rasmussen 2004 and Mark Schroeder 2010.
[316]   See Derek Parfit 2001 and Christian Pillar 2001.

bear) on a question the settling of which amounts to the forming of the attitude.[317] Hieronymi

develops her account in such a way that the only right kind of reason for belief is evidence. Her

thought is that the question the settling of which amounts to the forming of belief in *p* is simply

the question 'whether *p*?', and that the only considerations that bear (or are taken to bear) on this

question are pieces of evidence for and against *p*.

    I will assume that the epistemic/non-epistemic distinction is a special case of the right

kind/wrong kind distinction. However, I will not assume any particular account of the latter

distinction. But this is not to say that we have no way to think about what it is for a reason to be

epistemic. Instead, following Mark Schroeder,[318] we can see that there are certain data points or

earmarks on the basis of which the right kind/wrong kind distinction seemed like it needed to be

drawn in the first place, and to which any account of the distinction would need to be beholden,

to be ultimately plausible. What I will say – because it is the best I can do – is that I will assume

that if consideration aligns with these earmarks in the right way, it is a right kind of reason. I

leave the question of precisely how to draw the right kind/wrong kind distinction to others, or at

least to another day.

    There are three earmarks.[319] The first we have already met: the right kind of reason for an

attitude bears on a kind of rationality suited or appropriate to the nature of that attitude. The

second earmark is asymmetry in motivational efficacy. Wrong kind of reasons are hard, if not

impossible, to actually respond to, while right kind of reasons are the kind of considerations for

which one can form the attitude. Notice how hard it is to believe it is not raining for the bribe,

and how easy it is to believe on the basis of the weatherman's prediction.

---

[317]   See Hieronymi 2005 and 2013.
[318]   Schroeder 2012a.
[319]   Ibid., 458-461.

The third and final earmark concerns what we shall call *correctness*. As this earmark shall be a main concern in this chapter, I begin with some clarificatory remarks on what correctness amounts to. To the degree an attitude is correct, the attitude 'gets it right' in an objective sense; to the degree that it is incorrect, the attitude 'is defective' or 'gets it wrong' in an objective sense. For example, a belief, if true, gets it right to at least some degree, while a belief, if false, is defective to at least some degree. Likewise, admiring a kind, intelligent, hardworking person gets things right to at least some degree, while admiring a nasty person gets things wrong to at least some degree. On two fronts, correctness has a structural relationship to rationality. On the one hand, an attitude's being rationally formed and maintained is the agent's doing her part, in at least some sense,[320] towards making it the case that the attitude is correct. On the other hand, correctness and rationality can come as wide apart as we please. If fortune smiles on one, an attitude can have an arbitrarily high degree of correctness even if completely *ir*rational, as when one admires a person who appears nasty but is actually kind, intelligent, and hardworking. And if fortune frowns on one, an attitude can have an arbitrarily high degree of *in*correctness even if completely rational, as when one admires a person who appears kind, intelligent, and hardworking but is actually nasty.

With this clear, the third and final earmark of the right kind/wrong kind of reason distinction is asymmetric bearing on correctness. The right kind of reasons bear, in some way, on whether it is correct to form the attitude – as, for example, evidence bears on truth, which goes at least some way towards a belief's being correct. But the wrong kind of reasons do not bear on whether it is correct to form the attitude – as, for example, a bribe to admire a nasty person does not bear on whether it is correct to admire him.

---

[320]   For discussion of this general idea, see Wedgwood 2002b.

In this chapter, I will assume that a consideration possessing each of these three earmarks, in the case of belief, is an epistemic reason for belief. I will also assume that if a consideration lacks any of these earmarks, it is not an epistemic reason for belief. This last assumption will only make it harder for me to discharge my argument that some practical reasons are epistemic, but it clarifies the dialectic.

§9.3   Two Arguments for Evidentialism

Now we have a rough idea what it would mean to say some epistemic reasons are practical. But we do not yet know what it would take to show this. So, in this section, I relate two arguments for evidentialism. These arguments will help us see the problems we would need to solve, to plausibly maintain that some epistemic reasons are practical reasons.

§9.3.1 Thomas Kelly's Argument for Evidentialism

Thomas Kelly argues that only evidential reasons can bear on epistemic rationality. He begins with the point that realizing a belief in *p* has good consequences is not the kind of thing that can result, by itself, in believing *p*.[321] By contrast, realizing that *p* has evidence in its favor is the kind of thing that can result, by itself, in believing *p*. That is, it appears practical reasons are impotent to affect belief. But if a kind of consideration is impotent to affect belief, it cannot make a difference to whether the belief is rational. Kelly draws comparison to the benefits of being a certain height and the "rationality" of being a given height.[322] There may well be non-negligible benefits of not being two feet shorter than one in fact is, and one may be well aware of this fact. However, realizing this is not the kind of thing that can make one taller or shorter. If these considerations *were* able to affect one's height, they might well make a difference to whether one's height is rational. But they aren't, and so they don't. Similarly, since practical

321   Kelly 2002, 166.
322   Ibid., 166-167.

reasons are impotent to affect one's beliefs, they cannot make a difference to whether they are rational.

Of course, one can engage in activities that may increase the chances of exposure to evidence for the beliefs with good consequences, thus affecting which beliefs one forms. But

> Although practical considerations can make a difference to what one believes, they do not do so by constituting grounds on which beliefs are based….And rational beliefs, like rational actions, are rationalized by those considerations on which they are based.[323]

As it turns out, Kelly is drawing a connection between two of the earmarks of the right kind of reasons. In brief, his argument runs as follows. A reason bears on epistemic rationality only if it has motivational efficacy, that is, only if it is the kind of thing on which a belief can be based. But practical reasons are motivationally inefficacious. So, they do not bear on epistemic rationality and, hence, are not epistemic reasons. So, evidentialism is true.

Since I am assuming a reason lacking an earmark of a right kind of reason is not an epistemic reason, the premise of Kelly's argument that makes all the trouble is the premise that practical reasons are motivationally inefficacious. We can easily grant the other premises. But Kelly's argument does not, at least by itself, provide much support for the idea that practical reasons are motivationally inefficacious, instead relying on intuition in light of certain cases. So let us turn to another argument for evidentialism.

§9.3.2 Nishi Shah's Argument for Evidentialism[324]

Nishi Shah argues for evidentialism from *transparency*, which is the following thesis:

---

[323]   Ibid., 174.

[324]   Nishi Shah (2006, 482) defines evidentialism a bit differently than I do. He defines it as the view that only evidential reasons are reasons for belief, rather than the view that only evidential reasons are epistemic. But since epistemic reasons are characterized as appropriately related to the nature of belief, Shah's arguments can also be understood as supporting evidentialism as I understand it. Notice that evidentialism is often contrasted with externalist views, like Plantinga's (1993) proper functionalism, on which evidence is not required for justified belief. In this chapter, evidentialism is instead contrasted with views on which it is not just evidential reasons that are epistemic.

> …the deliberative question *whether to believe that p* inevitably gives way to the factual question *whether p*, because the answer to the latter question will determine the answer to the former.[325]

The first question gives way to the second in the sense that one immediately takes it to be

"settled by, and only by, answering" the second question, and deliberates accordingly.[326] The

sense in which this is inevitable is that it is an "unalterable psychological fact".[327]

Transparency is a purported observation about *doxastic deliberation*.[328] Not just any

mental event in which beliefs are formed or revised is an instance of doxastic deliberation.

Doxastic deliberation is guided by an intention to form a belief with respect to *p*.[329] Thus, it

involves entertaining of the concept of belief.[330] Moreover, transparency is a purported *first*

personal observation about doxastic deliberation: the claim is we find that the one question

collapses into the other when we engage in such deliberation.[331]

Shah adds to transparency *the deliberative constraint on reasons*. This constraint is the

view that *x* is an epistemic reason only if *x* is able to dispose the subject S towards or against

believing *p* "in the way characteristic" of *x*'s "functioning as a premise in doxastic

deliberation".[332] We can, more simply, understand this as the constraint that an epistemic reason

must be motivationally efficacious in doxastic deliberation.

Evidentialism follows from transparency and the deliberative constraint on reasons.

Given transparency: in doxastic deliberation, it is an unalterable psychological fact that the

question *whether to believe p* gives way to the question *whether p*. But the only considerations

---

[325]   Ibid., 481-482.
[326]   Shah 2003, 447.
[327]   Shah 2006, 484.
[328]   Ibid., 481.
[329]   Ibid., 482.
[330]   Ibid., 490 and Shah 2003, 471.
[331]   Shah 2003, 447.
[332]   Shah 2006, 486. Actually, Shah defines this constraint in terms of reasons for belief rather than in terms of epistemic reasons. But this difference will not matter; see footnote 324. It is worth noting that, according to Shah's (Ibid., 484) interpretation, Bernard Williams (1979) gives an argument for this constraint to the effect that a reason for S to do something must be something for which S can do it, and that a reason for which S can do something must be something S can treat as a reason to do it.

bearing on the latter question are pieces of evidence for and against *p*. So, transparency implies that only evidential reasons are motivationally efficacious in doxastic deliberation. But given the deliberative constraint on reasons, only reasons motivationally efficacious in doxastic deliberation could be epistemic reasons. So, evidentialism is true: all epistemic reasons are evidential.

Before developing a reply to Shah's argument, it is worthwhile having on the table his explanation of transparency.[333] The reason why *whether to believe p* collapses into *whether p* in doxastic deliberation is that doxastic deliberation is an activity governed by a particular norm. That is, one counts as engaged in this activity only if, in the first place, one is aiming at a result in accordance with the norm of correctness for belief. And that norm is that a belief in *p* is correct if and only if *p* is true. Moreover, generally, reasons bearing on an activity governed by a norm of correctness are nothing other than considerations indicating whether a given result of the activity is correct, considerations for which one is able to terminate the activity with that result as guided by the norm of correctness. Reasons indicating whether a given result of doxastic deliberation is correct are all and only evidence, given the norm of correctness for belief.

One possible reaction to Shah's argument is to deny that epistemic reasons need to be motivationally efficacious in doxastic deliberation in particular and admit that, in doxastic deliberation, only evidence has motivational efficacy. This would be to reject the deliberative constraint on reasons, which says that epistemic reasons must be motivationally efficacious in doxastic deliberation, while accepting transparency, on which *whether to believe p* collapses into *whether p* in doxastic deliberation. Notice that it is *not* to deny that epistemic reasons need to

---

[333]   Shah 2006, 489.

have motivational efficacy, because one can hold that they have such efficacy in other contexts – contexts in which one is *not* doxastically deliberating.

But if Shah's is the right explanation of transparency, there is a worry for this reaction to Shah's argument. On Shah's explanation of transparency, a belief is correct just in case it is true, and that is the end of the story about the conditions for correctness of belief. But if this is right, then only evidence bears on correctness, since only evidence bears on truth. And an earmark of a right kind of reason is that it bears on correctness. So, on Shah's explanation, only evidential reasons are epistemic, once again.

§9.4   Responding to the Arguments for Evidentialism

Before developing a response, it is worthwhile clarifying how we should think about Kelly and Shah's arguments given our assumptions. Remember, in §9.2, we assumed that a reason is epistemic just in case it has the following three earmarks: motivational efficacy, bearing on rationality, and bearing on correctness.

§9.4.1 How to Think About the Arguments

I begin with Shah's argument, roping in Kelly's argument later. The first thing to notice about Shah's argument is that both premises – the deliberative constraint on reasons and transparency – are explicitly cast in terms of *doxastic deliberation*. The deliberative constraint on reasons says epistemic reasons must be motivationally efficacious in doxastic deliberation. And transparency says *whether to believe p* collapses into *whether p* in doxastic deliberation, which implies that only evidential reasons are motivationally efficacious in doxastic deliberation.

Now, according to Shah, doxastic deliberation about whether to believe p is always guided by an intention to form a belief with respect to p – that is, by an intention to either believe p or believe p's negation. This makes an argument cast in terms of doxastic deliberation rather

narrow in scope. For it does not take much reflection to notice that it is quite rare to intend to either believe or deny a particular proposition via some course of deliberation. Perhaps if you are a member of a jury, you will intend to either believe or deny the proposition that the defendant is guilty via the trial's proceedings, discussion with other jurors, examination of the evidence, etc. But I suspect it is not uncommon to go entire days if not weeks without ever explicitly intending to either believe or deny any particular proposition. That is, it is not uncommon to go entire days if not weeks without engaging in doxastic deliberation.

Despite this, we form, revise, affirm, and abandon beliefs by the hour if not the minute. Moreover, in some cases we do this *rationally* and in other cases we do not. But a view about epistemic reasons should be a view about the reasons we respond to rationally (or not) in all contexts, including these quotidian ones, and not just in the rare cases of doxastic deliberation.

But we can understand Shah's argument in a more general way. If an epistemic reason must be motivationally efficacious in doxastic deliberation, presumably this is precisely because an epistemic reason must be motivationally efficacious *in general*. So, we can let our assumption that epistemic reasons must be motivationally efficacious replace the deliberative constraint on reasons. Indeed, our assumption is more general, and implies the deliberative constraint on reasons.

In addition, on our assumptions, we can spot Shah the resources for a premise that is more general than transparency. Shah casts transparency in terms of doxastic deliberation for the following reason. According to Shah, when one intends to form a belief with respect to *p* and so entertains the concept of belief, one's activity is governed by the norms for correctness on belief. And according to Shah, there is only *one* such norm:

the Truth Norm   It is correct to believe *p* if and only if *p* is true.

The view that the Truth Norm is the only norm of correctness on belief is a premise more general than transparency. If this premise is correct, then absolutely generally, *whether to believe p* collapses into *whether p*. And if that is right, then only evidential reasons have motivational efficacy.

Shah's argument gives us reason to think only evidential reasons have motivational efficacy. Enter Kelly. Kelly argues that only motivational efficacious reasons bear on epistemic rationality, and that only reasons bearing on epistemic rationality are epistemic reasons. So, when both arguments are combined, evidentialism is the result: only evidential reasons are epistemic reasons.

So, we can combine Shah and Kelly's arguments into a single line of thought:

(v)   The only norm of correctness on belief is the Truth Norm.

(vi)   If the only norm of correctness on belief is the Truth Norm, then only evidential reasons have motivational efficacy.

(vii)   If only evidential reasons have motivational efficacy, then only evidential reasons bear on epistemic rationality.

(viii) If only evidential reasons bear on epistemic rationality, then only evidential reasons are epistemic reasons.

(ix)   Only evidential reasons are epistemic reasons.   [from (v), (vi), (vii), and (viii)]

(ix) is evidentialism. Given our assumptions, this argument indulges in overkill. For on our assumptions, epistemic reasons have motivational efficacy; so, (v) and (vi) alone suffice to establish (ix). But the whole argument is worth having on the table, because it is a testament that others take seriously the three earmarks of epistemic reasons: motivational efficacy, bearing on rationality, and bearing on correctness.

Now given the three earmarks on epistemic reasons, (vii) and (viii) are true. Moreover, (vi) is quite plausible on its face. In some sense, a belief's internal or constitutive aim is to be

correct, and what could easily motivate belief unless it is connected to this aim? But only evidential reasons are connected to whether a belief is true. So, if the only norm on correctness is the Truth Norm, only evidential reasons have motivational efficacy, just as (vi) says. If this is right, then what we learn from the Kelly-Shah argument is that whether evidentialism is true turns on whether the only norm of correctness is the Truth Norm. So, let us turn our attention to the Truth Norm.

§9.4.2 The Inadequacy of the Truth Norm

A few preliminary, clarificatory comments are in order. Given my argument in the last chapter that evidence pushes our credences around, it might well be that only evidential reasons are motivational efficacious for credences. In light of this, I am willing to accept that the only norm of correctness on *credence* is the truth norm on credence, namely that it is correct to have a credence of n in p to the degree n approaches 1 if p is true and 0 if p is false.[334] And if the truth norm on credence is the only norm of correctness on credence, this may partially explain why it may seem to be intuitive that the Truth Norm is the only norm of correctness on belief. For credence and belief are not always appropriately distinguished.

Even so, I shall argue the Truth Norm is not the only norm of correctness on belief. It is not adequate, all by itself. Remember that belief is something like a defeasible disposition to treat a proposition as true in one's reasoning. In fact, in Chapter 8, I developed the idea that belief is the state one is in by virtue of organizing one's attention as if p is true, that is, by virtue of attending only to possibilities consistent with p. But as I shall argue, if belief is like this, the Truth Norm alone is inadequate to guide us. In fact, we can show that the Truth Norm is inadequate to guide us even within doxastic deliberation. Doxastic deliberation is, so to speak,

---

[334]   See Wedgwood 2013b, 226-227.

Shah's home court. If the Truth Norm is inadequate to guide us in doxastic deliberation, then, *a fortiori*, it is inadequate to guide us generally.

Doxastic deliberation about whether to believe *p* begins with an intention to either form belief in *p* or form belief in *p*'s negation, and comes to a successful end when this is done. When one has this intention, the norms of correctness on belief guides one. But if the Truth Norm is the only such norm, then the question *whether to believe p* collapses into the question *whether p*, and one looks for evidence relevant to *p*. This part of Shah's argument for evidentialism we can accept.

Now suppose that, during doxastic deliberation, one finds evidence on which *p* is quite likely. Has one answered the question *whether p*? No. Being quite likely on one's evidence and being true can come apart. So, finding that *p* is quite likely is not to answer whether *p*. One thing we do know, though, is that whether the question *whether p* has been answered looks to be a function of how probable (improbable) *p* is on the evidence uncovered. Once we recognize this, though, it is clear that the only function from a probability of *p* to an answer to *whether p* is a function mapping a probability of 1 to an affirmative answer and a probability of 0 to a negative answer, where all other values yield no answer at all. We can see this by thinking about a proposition concerning a given lottery ticket, to the effect that that ticket will lose. Our evidence can make the probability of that proposition arbitrarily high without yet giving us an answer to the question of whether the ticket will lose. What the lottery case shows is that evidence always leaves the question of *whether p* unresolved, unless indicating a probability of 1 or 0.

If the Truth Norm is the only norm of correctness on belief, then *whether to believe p* collapses into *whether p*, and evidence alone guides us in doxastic deliberation. But, as we have seen, if evidence alone guides us, our question is unresolved unless and until the evidence we find gives *p* a probability of 0 or a probability of 1. And this shows that the Truth Norm is

inadequate to guide us. Regularly and without compunction we answer the question *whether to believe p* without finding evidence on which the probability of *p* is either 0 or 1. In particular, regularly and without compunction we form something like a defeasible disposition to treat a proposition as true in one's reasoning. Even more specifically, in light of Chapter 8, regularly and without compunction we organize our attention as if *p* is true, and by virtue of beginning this activity enter into the state of belief. What this shows is that the Truth Norm is *inadequate* in doxastic deliberation. It cannot be the only norm of correctness that governs or guides the activity of doxastic deliberation. *A fortiori*, it cannot be the only norm of correctness that governs belief formation, revision, etc., *generally*.

§9.4.3 Practical Norms of Correctness

In light of the above, it is plausible that the Truth Norm is not the only norm of correctness on belief. Even so, we can accept that the Truth Norm is true. All we need to do is hold that correctness is *degreed*. If we do, then we can also hold that belief in *p* is correct if *p* is true and incorrect if *p* is false, but its degree of correctness or incorrectness depends in some way on the practical facts. This view is true just in case the Truth Norm is true and there is at least one practical norm of correctness on belief. And in light of the last section, it is the kind of view we should expect, given the inadequacy of the Truth Norm if left by itself.

There is independent reason to suspect there are practical norms of correctness on belief. Consider the fact that sometimes it is more important to have a true belief than to avoid having a false one while, other times, the situation is reversed. For the first kind of case, suppose that you are the captain of a ship. Suppose that you have evidence on which it is quite likely, but of course not absolutely certain, that there is a dangerous reef at location X. For the second kind of case, suppose that you have evidence on which it is quite likely, but of course not absolutely certain, that the firearm on the table is unloaded. Suppose that you believe there is a reef at X,

and also believe that the firearm is unloaded. If these beliefs are true, then each has a degree of correctness. Even so, the true belief that there is a reef at X seems objectively better than the true belief that the firearm is unloaded. It is constitutive of true belief to be good, but some true beliefs are better than others, and better not accidentally or instrumentally but instead precisely on account of their playing their normal, characteristic role in our mental economy. So, plausibly, the true belief that there is a reef at X has more correctness than the true belief that the firearm is unloaded. Now if these beliefs are false, then each has a degree of incorrectness. Even so, the false belief that the firearm is unloaded seems objectively worse than the false belief that there is a reef at X. It is constitutive of false belief to be bad, but some false beliefs are worse than others, and worse not accidentally or instrumentally but instead precisely on account of their playing their normal, characteristic role in our mental economy. So, plausibly, the false belief that the firearm is unloaded has more incorrectness than the false belief that there is a reef at X.

If the above is on track, then there are practical norms of correctness on belief on account of the fact that the importance of having a true belief as against avoiding a false belief can vary depending on the practical facts. There may be other kinds of practical norms, too. For example, it might be that beliefs in true trivialities are not as correct as beliefs in true substantive propositions. It is not my goal, though, to delineate the possible kinds of practical norms of correctness on belief.

That being said, we need not think norms of correctness on belief are a disorganized hodgepodge. Reflect on the nature of belief: belief in p is something like being defeasibly disposed to treat p as true in one's reasoning, namely, it is the state one is in by virtue of organizing one's attention as if p is true. In this way, to believe p is for p to be part of one's picture of the world, the picture on which one relies. D.M. Armstrong said that one's set of

beliefs is a map that represents the world and "*by which we steer*".[335] To think of it metaphorically, to believe *p* is for *p* to be part of one's map. With our credences, we can reason about the likelihood of the presence and location of items that are nowhere represented on the map. But we need a map if we want to get anywhere, because we need to take some things for granted to get anywhere.

The map metaphor may yield dividends. What features of a map goes towards its being an excellent map in an objective sense, and which go towards its being defective? Well, only accurate elements of the map go towards its being excellent, and inaccurate elements always go towards its being defective. But notice the following. Compare a map which represents the reef at location X, but is silent about whether the gun is unloaded, with a map which represents the gun as unloaded, but is silent about the reef. The first map is more excellent than the second. For part of what goes towards excellence is accurately representing what is important to accurately represent. We might also compare a map otherwise identical to another save where the other has a handful of trivial representations, it has a handful of substantive ones. The map with more relevant, substantive information is more excellent than the other.

The kinds of things which go towards making a map excellent (defective) appear to also go towards making a set of beliefs excellent (defective). Having information when a lack of information would be disastrous (consider the reef as compared to the unloaded gun), having non-trivial information, etc., all go towards a map being excellent. But these features also go towards a set of beliefs being excellent. And just as these features go beyond mere accuracy in the case of a map, they go beyond mere truth in the case of a set of beliefs. These features constitute what Alvin Plantinga calls epistemic values other than truth:

---

[335]   Armstrong 1973, 3.

> …something must be said about *other* epistemic values here [besides truth]: the importance of considering *important* propositions, of having beliefs on certain crucial topics, of avoiding unnecessary clutter and frivolous dilettantism, perhaps of having a coherent set of beliefs, and so on.[336]

Now, though I have not offered a list of these other values, we need not view norms of correctness on belief as a disorganized hodgepodge. For we can organize all of these norms with the following idea: a belief is correct to the degree it constitutively contributes to the excellence of one's set of beliefs. The contribution has to be *constitutive*, of course: we need to screen off cases in which a demon will give us many true beliefs in return for a single false one, and etc. But with this clear, any true belief makes some contribution to the excellence of one's set of beliefs, while any false belief detracts from it. However, some true beliefs make greater contributions, just as some accurate elements of a map make greater contributions to a map's excellence, and *mutatis mutandis* with respect to false beliefs.

Understood in this way,

the Excellence Norm A belief in *p* is correct to the degree that it constitutively contributes to the excellence of one's set of beliefs, and incorrect to the degree that it constitutively detracts from the excellence of one's set of beliefs.

not only unifies the norms of correctness on belief, it also vindicates that

the Truth Norm Has Company A belief in *p* is correct if true and incorrect if false, but its degree of correctness or incorrectness depends in some way on the practical facts.

It is good that the Truth Norm Has Company. Remember that the Truth Norm, by itself, is inadequate. It is inadequate to guide us when the question is whether to believe *p*, but we do not have evidence on which *p* is certain. But on the present picture, the Excellence Norm can guide us, for example, to believe that the reef is at location X, but to withhold on whether the

---

[336]   Plantinga 1993, 33.

firearm is unloaded. It is not just evidence which bears on correctness, given the Excellence Norm.

Now, I have been speaking of how different reasons "bear on" correctness given different accounts of correctness. I have not explained how to understand "bear on". One option is to say that reasons of which we are conscious bear on correctness insofar as they somehow make available to our consciousness the *expected correctness value* of belief, as against suspension. If the Truth Norm is left by itself, the belief that the reef is at location X and the belief that the firearm is unloaded have the same expected correctness value, namely some positive value exactly corresponding to one's evidence. But given the Excellence Norm, the belief that the firearm is unloaded may actually have *negative* expected correctness value, given the high degree of incorrectness the belief would have if false and the low degree of correctness it would have if true. On this account, we can see how the Excellence Norm could guide one to believe the reef is at location X, but to withhold on whether the firearm is unloaded. Now, I see no need to commit myself to this particular account of "bearing on". Perhaps other accounts could also do the job. I mention it only to illustrate in more detail how it might be that the Excellence Norm, by contrast with merely the Truth Norm, is adequate to guide us.

So, the Truth Norm Has Company. There are practical norms of correctness on belief. Thus, premise (v) of the Kelly-Shah argument is false. Moreover, now we have reason to expect that there are practical epistemic reasons, if we only turn around the reasoning behind the Kelly-Shah argument for our own purposes. If the Truth Norm Has Company, then some practical reasons bear on correctness. If some practical reasons bear on correctness, then some practical reasons have motivational efficacy. If some practical reasons have motivational efficacy, then some practical epistemic reasons bear on epistemic rationality. So, some practical reasons have each of the three earmarks of the right kind of reasons. So, some practical reasons are epistemic.

This reason to expect there are practical epistemic reasons dovetails nicely with a reason to expect as much which I mentioned at the end of Chapter 8. On my account, belief is a state grounded in an activity. It is plausible that practical reasons and activities are, as it were, made for each other. So, on my account of belief, we should expect that some practical reasons are the right kind of reasons for belief.

§9.5   On Behalf of Practical Epistemic Reasons

If my argument is on track thus far, practical reasons have at least *one* of the earmarks of epistemic reasons: they bear on the correctness of the belief. As I have explained, we should expect them to have the other two earmarks. Now it is time to look and see.

Mark Schroeder has argued that there are practical epistemic reasons by giving cases in which, intuitively, certain practical reasons are both motivationally efficacious and bear on epistemic rationality.[337] I shall relate Schroeder's argument in terms of one example he uses, which I call the *Biopsy Case*.[338] I shall defend Schroeder's argument, and discuss how the Excellence Norm helps clarify matters.

In what I will call *the Biopsy Case,* you have many times in the past had suspicious looking skin spots checked for cancer. Each time the spot has proved to be benign rather than cancerous, so you have strong inductive evidence that this particular time your skin spot will also be benign. However, the lab has promised to call tomorrow evening and give you the results of its tests. That the lab will call later is not evidence that your spot is benign, and it is not evidence that it is not benign, but it does seem to be a reason to not believe that it is benign, despite your having strong inductive evidence that it is benign. It is also a reason to not believe that your skin spot is not benign. Intuitively, it is a reason, more simply, to suspend judgment about whether it

---

[337]   See Schroeder 2012a.
[338]   Ibid., 471-472.

is benign. But, again, it is not an evidential reason. So, it is a practical reason to suspend. Put

differently, *p* could only be a practical reason, vis-à-vis adopting belief, denial, or suspension

towards *q*:

> *p*   The lab will call tomorrow evening and give you the results of its tests.
>
> *q*   Your skin spot is benign.

And here is the question: is *p* an epistemic reason to suspend judgment on *q*? Well, to

begin, it appears easy for you to suspend belief with respect to *q* for or on the basis of *p*.

Suspending on the basis of *p* is nothing like suspending belief for the reason that you have been

bribed to suspend, which would be much more difficult to do. Moreover, it seems epistemically

rational to suspend on the basis of *p*, not rational in the sense one would be rational in

suspending for a bribe.

The Excellence Norm helps explain these facts. Given this norm, *p* can bear on the

degree of correctness belief in *q* has if true and the degree of incorrectness belief in *q* has if false.

As on the construal of "bearing on" I entertained above, this might work as follows: on the

reasons of which one is conscious, the expected correctness value of belief in *q* may be negative,

given the high degree of incorrectness a belief in *q* has if false and low degree of correctness a

belief in *q* has if true. Being misinformed about *q* might make one's life go much more poorly

than it otherwise could have, given that conclusive evidence regarding *q* is soon forthcoming. On

the other hand, there is very little benefit to believing *q* preemptively, before the lab calls. Even

if a map including *q* is more accurate and steering one's life by this map allows one to go to the

movies tomorrow evening carefree, there is too much to lose in missing the call from the lab. By

contrast, if no evidence regarding *q* will ever be forthcoming, transitioning from merely a high

credence in *p* to outright belief does give one a much more excellent map and, plausibly, would

not make one's life go more poorly than it otherwise could have even if, by chance, *q* turned out

to be false. In this way, given the Excellence Norm, *p* bears on degrees of correctness, which would lead us to expect *p* to have motivational efficacy and bear on rationality. And this is exactly what we find, here.

In any case, if this line of thought is correct, then some practical reasons have each of the three earmarks of epistemic reasons and so, on our assumptions, qualify as epistemic reasons.

But there is an important objection to this argument on behalf of practical epistemic reasons. Several have argued that cases like the Biopsy Case conflate reasons for outright doxastic states, at a crucial point, with reasons for and against deliberating about what to believe.[339] That the lab will call with the results is a reason, one might think, for you to hold off deliberating about whether to believe your skin spot is benign. But it is not a reason to *suspend*, or to *not believe*, or etc. So it is not a reason for a doxastic state but instead a reason to not deliberate. And, so, it is not an epistemic reason.

But Schroeder's reply to this objection is, I think, convincing.[340] Begin by considering two variations of the Biopsy Case. In the first variation, you have already inferred from the inductive evidence that your skin spot is benign before you are told that conclusive evidence from the lab will arrive tomorrow evening. In the second variation, you have not thus inferred, you have no belief either way when you are told that the lab will call. If the fact that the lab will call is really a reason to *not deliberate*, then it is a reason to not deliberate in each of these variations of the case. But that is false. If you already believe that your skin spot is benign, that the lab will call soon is a reason to *stop* believing this, to transition from belief to suspension, at least for the time being. We wrongly think the reason is a reason to not deliberate because we

---

[339] For this objection, see Shah and Silverstein (2013), Graham Hubbs (2013), and Hieronymi (2013). Unlike in the case of Shah's argument, there is nothing meaty packed into the notion of 'deliberation', here.

[340] See Schroeder (2013, 134-136), though the reply is cast in terms of intentions, it is clearly meant to apply also to the case of belief, which is how the reply is understood in the main text.

had imagined that you have no belief either way, because then the status quo is to not believe, and not deliberating always preserves the status quo. But if we consider the variation where the status quo is different, we see that the reason at issue is not a reason to not deliberate. Instead, we see that it is a reason against believing and for suspending, directly.

Besides, if the Excellence Norm is correct, we should expect the fact that the lab will call to bear on the degree of correctness of belief. And so, as I have explained, we also should expect this fact to have the other earmarks of epistemic reasons.

And finally, though Schroeder does not go so far as to say this, if we imagine a variation of the Biopsy Case in which it somehow becomes clear that no evidence about whether your skin spot is benign will ever be forthcoming, I propose that this fact is a practical reason to *not* suspend. Put differently, it is a practical reason to either believe or not believe. In this kind of case, in which you already have solid inductive evidence that your skin spot is benign, not suspending will presumably constitute believing that your skin spot is benign. So, this kind of practical reason may, in this kind of case, make it rationally permissible for you to believe your skin spot is benign. For holding this belief gives you a more useful map by which to steer your life, and in the absence of any further evidence available, there is very little to lose by doing so.

In any case, so concludes my argument against evidentialism, and on behalf of the view that some epistemic reasons are practical.

§9.6   Conclusion

It is worthwhile looking one last time at

The At the Mercy of Evidence Objection, Specific Form

(i)   If an obligation is epistemic, it is an obligation to respond to one's evidence in a certain way.

(ii)   If one has the relevant kind of obligation to respond to something in a certain way, one's response to it is an exercise of will.

(iii)   A human person's response to evidence is never an exercise of will.

(iv)   Therefore, no human person ever has an epistemic obligation.
[from (i), (ii), and (iii)]

In this chapter, I have argued that (i) is false. There are more than just evidential epistemic reasons, and so an epistemic obligation is not simply an obligation to respond to one's evidence in a certain way.

So much for the At the Mercy of Evidence Objection. In the next and final chapter, I deal with another objection to deontological explanations of accessibilism. In responding to the At the Mercy of Evidence Objection, the distinction between credence and outright belief has proved important. This same distinction is the key to solving the objection I address in the next chapter.

Chapter 10: Dissolving the Dilemma

§10.1   Introduction

In Chapters 7-9, I provided a response to the At the Mercy of Evidence Objection. In the process, we have developed in more detail the deontological explanation of accessibilism first offered in Chapters 4 and 5. In doing so, we have developed not only an explanation of accessibilism, but a full-fledged accessibilist theory. As we shall see, one component of this theory is a special instance of the view that awareness of reasons is required for rationality. Several have offered a dilemma for this kind of view. They argue that, however awareness is understood, either the view that awareness is required is unmotivated, or the view generates problems for foundationalism. This objection is what I called *the Dilemma* in §1.3.2. I shall show that my account has the resources to reply to the Dilemma.

The chapter is structured as follows. In §10.2, I summarize the accessibilist theory to which we have arrived, and explain how it involves a special instance of the view that awareness of reasons is required for rationality. I call this special instance *the Consciousness Requirement*. In §10.3, I relate the versions of the Dilemma offered by Laurence BonJour and Michael Bergmann. In §10.4, I explain how the Dilemma applies to the Consciousness Requirement. I show that our theory has the resources to reply to the Dilemma in §10.5. I conclude in §10.6, with some final remarks on what to think about deontologism, for which I have not yet argued in this dissertation.

§10.2   The Accessibilist Theory

In the preceding chapters, I developed the following explanation of accessibilism:

deontologism   It is rational for S to respond to epistemic reasons in a given way at *t* if and only if it is blameless for S to respond in that way at *t*.

The Bridge Principle	Whether it is blameless for S to respond to epistemic reasons in a given way at t is determined by the facts about which things are accessible to S at t.
accessibilism	Whether it is rational for S to respond to epistemic reasons in a given way at t is determined by the facts about which things are accessible to S at t.

Moreover, the way this explanation was developed informs what accessibilism *amounts to*.

I begin with some comments on deontologism. To say it is blameless for S to respond to epistemic reasons in a particular way is to say that doing so fulfills an obligation S has at the time. An obligation applies to S only if there is both something blameworthy S can do at the time as well as something blameless S can do at the time instead. So, it is blameless for S to respond to epistemic reasons in a particular way only if S can respond to these reasons in a blameworthy way at the time. And notice that this blameworthy response must be available *at the time*. Blameworthiness is understood here as *original* rather than *derivative*: this property always attaches to what one does at the time, and whether one has this property at the time can never derive from blameworthy things one did in the past.

Blamelessness is the avoiding of presently available blameworthiness. In Chapter 4, I argued that blameworthiness requires the person herself to exercise control, and that this requires consciousness of the reasons to which she responds. In this way, I argued that blameworthiness is presently available only if one is *conscious* of a particular set of reasons. So, since blamelessness is the avoiding of presently available blameworthiness, it is blameless for one to respond in a particular way to reasons only if one is conscious of those reasons. Moreover, whether it is blameless for one to respond in a particular way to reasons of which one is conscious is determined by the facts about which things are accessible to one; so, accessibilism follows.

In Chapter 5, I identified the person's control with *the will*. The will is a capacity which, if exercised, involves the experience of oneself as the source of one among multiple available alternatives, and which is exclusively responsive to reasons of which one is conscious. This led us to, in Chapters 6-9, establish that this deontological explanation is to be cast in terms of *belief* rather than *credence*. For, in these chapters, we saw that no exercise of will is the forming of a particular credence while, on the other hand, it is possible for an exercise of will to be the forming of a belief.

The result is a full-fledged accessibilist theory. On this theory, the question of whether, in the relevant sense of 'rational', it is rational for S to form a belief in *p* arises only when S is conscious of epistemic reasons to which forming belief in *p* is one of multiple responses available to S at the time, where at least one of the other responses would be blameworthy. And whether it is rational for S to form the belief is determined by the facts about which epistemic reasons S is conscious of at the time. These facts determine *propositional* rationality. Whether S is *doxastically* rational is a matter of how S exercises S's will in response to those epistemic reasons of which S is conscious. That is the theory.

The following is a component of that theory:

The Consciousness Requirement	It is rational for S to form belief in *p* at *t* only if S, at *t* is conscious of reasons in response to which S can do so.

The Consciousness Requirement is a special instance of a more general kind of view, which we shall call

The Awareness Requirement	Epistemic justification or rationality requires awareness of reasons.

A certain kind of dilemma has often been pressed against the Awareness Requirement. Since the Consciousness Requirement is a special case of the Awareness Requirement, the Consciousness

Requirement is in trouble if this dilemma has force. And since the Consciousness Requirement is a built-in component of our accessibilist theory, that theory is also in trouble if this dilemma has force. I turn to instances of this dilemma in the next section.

§10.3   The Dilemma for the Awareness Requirement

The dilemma for the Awareness Requirement follows a certain procedure. First, awareness is split into two mutually exclusive and jointly exhaustive kinds. Then it is argued that, depending on which kind of awareness is at issue in the Awareness Requirement, the view either clashes with foundationalism or loses its motivation.

§10.3.1   *Foundationalism*

Before diving into instances of the dilemma, I want to make clear both what I shall take foundationalism to be and why preserving foundationalism is a desideratum. I should flag from the start, though, that I shall understand both foundationalism and the dilemmas we shall meet in terms of the *propositional* rationality of *events* of forming, revising, etc., (outright) *beliefs*. The philosophers I discuss are sometimes non-committal on each of these three points. But it should be apparent that the dilemmas they have in mind have a viable form or version in these terms. And it is versions of their dilemmas in these terms that, later, we can lean against the Consciousness Requirement.

With this in mind, I shall understand *foundationalism* to be the view that there are cases in which it is rational to form a belief, but it is *not* the case that it is rational to do this by virtue of facts about which other beliefs one has at the time.[341] On foundationalism, it is possible for *none* of the facts by virtue of which it is rational to form a belief to refer to other beliefs. Now, it

---

[341]   As Matthias Steup (2005) put it in his SEP article, according to foundationalism, there are cases in which "S's belief that *p* is justified without owing its justification to any of S's other beliefs". Notice that his definition does not distinguish propositional from doxastic justification, or the mere having of a belief from the forming of one.

may yet be that one responds to *reasons* in these cases. But these reasons will not be the objects of beliefs.

Let me explain why I think it is important to preserve foundationalism. If foundationalism is false, then it is rational to form belief in *p* at *t* only if one has other beliefs bearing some relation to *p* at *t*. For each of these other beliefs, it was either rationally formed (or at least rationally reaffirmed)[342], irrationally formed, or non-rationally formed (i.e., formed in such a way that the question of rationality does not arise). To the extent that these other beliefs were irrationally formed, the explanation for why it is rational for one to form belief in *p* is that one is, in a way, riding a wave of past irrationality. But it would be disconcerting if the *only* way for it to be rational to form belief is to, at least in some measure, ride a wave of past irrationality. It is puzzling why past irrationality should be an essential component of why present rationality is possible.

But, by the same token, it would also be disconcerting if the only way for it to be rational to form belief is to, at least in some measure, ride waves of past irrationality and/or waves of past non-rational events of belief formation. For believing is always susceptible to critique, to the kind of mental processes in which the question of whether it is rational to continue believing arises. And it would be nice if it were possible to critique non-rationally formed beliefs without the process of critique being, ultimately, an outgrowth of irrationally and/or non-rationally formed beliefs. In this way, at least, it would be nice if we can escape the circle of beliefs. For this reason, it would be nice if, in some cases in which it is rational to form belief, it is only beliefs rationally formed in the past that explain why it is rational.

---

[342]   A qualification I shall leave implicit for brevity, hereafter.

But given foundationalism's falsity, a vicious regress is generated from the supposition that its being rational to form belief in *p* is an outgrowth of *only* beliefs rationally formed in the past. This is because, for each rationally formed belief that bears some relation to *p* and helps explain why it is rational to form belief in *p*, there are *other* beliefs rationally formed in the past. And so on, *ad infinitum*. The steps in this regress are steps backward in time. And so it is vicious. Humans are not super-taskers: we are unable to complete infinitely many distinct events of belief formation in a finite amount of time. Therefore, denying foundationalism either leads to vicious regress, or else to the disconcerting result that present rationality is *always*, as it were, the crest of a wave of past irrationality formed beliefs and/or past non-rationally formed beliefs. If we accept foundationalism, though, we can hold that in some cases, what explains why it is rational to form belief in *p* is contact with something outside of the circle of beliefs.

These remarks are, admittedly, merely suggestive. A genuine defense of foundationalism, though, is beyond the scope of this dissertation. I shall simply assume that foundationalism would be nice to preserve, for the reasons discussed here.

§10.3.2   *BonJour and Bergmann*

Laurence BonJour and Michael Bergmann each argue that, depending on which kind of awareness is at issue in the Awareness Requirement, the view either clashes with foundationalism or loses its motivation.

BonJour takes epistemic rationality in forming a belief to be epistemic *responsibility* in forming the belief.[343] And he holds that one is epistemically responsible only if one is "in cognitive possession" of "a reason why it [the belief] is likely to be true".[344] If we add foundationalism to the mix, then it follows that in some cases of rational belief formation, one

---

[343]   BonJour 1985, 8.
[344]   Ibid., 32.

has a kind of awareness ("cognitive possession") of a reason that does not involve belief. This implication is what BonJour calls "the doctrine of the given".

He has a dilemma for this doctrine:

> …the proponent of the given is caught in a fundamental and inescapable dilemma: if his intuitions or direct awarenesses or immediate apprehensions are construed as cognitive, or at least quasi-judgmental (as seems clearly the more natural interpretation), then they will be both capable of providing justification for other cognitive states and in need of it themselves; but if they are construed as noncognitive, nonjudgmental, then while they will not themselves need justification, they will also be incapable of giving it. In either case, such states will be incapable of serving as an adequate foundation for knowledge. This, at bottom, is why empirical givenness is a myth.[345]

Awareness is either judgmental or nonjudgmental. If awareness is judgmental, it affects whether it is responsible for the subject to form the belief and hence is capable of "providing justification". But judgmental awareness is a kind of mental state of which we can ask whether the subject formed it responsibly. And since responsibility requires awareness of reasons, we need to posit *other* states of judgmental awareness of reasons. Thus if awareness is judgmental, foundationalism is in trouble, for the regress brewing here is precisely the kind of regress foundationalism is meant to stop.

Suppose, instead, that awareness is nonjudgmental. But then, worries BonJour, awareness is too thin and etiolated to affect whether it is responsible for the subject to form the belief. But then, given that rationality is responsibility, there is no reason to suppose, in the first place, that rationality requires awareness of reasons. So: if awareness is judgmental, foundationalism is in trouble; if nonjudgmental, the Awareness Requirement is unmotivated.

BonJour is less than fully explicit about the nature of the judgmental/nonjudgmental distinction, and about why nonjudgmental awareness cannot affect responsibility. However, his

---

[345]   Ibid., 69.

dilemma for the Awareness Requirement reflects a kind of worry philosophers are prone to have. So, rather than spend time reconstructing BonJour's dilemma, let us turn to a more explicit and developed instance of this kind of dilemma in the literature, namely Bergmann's dilemma.

Bergmann splits awareness into *strong* and *weak*.[346] *Strong* awareness of a reason is awareness of the reason that involves conceiving it as in some way relevant to the truth or appropriateness of the belief one may form. *Weak* awareness of a reason is awareness that is not strong.

Now suppose that the relevant awareness of reasons is strong. Then this awareness, since involving the conceiving of the reasons as relevant to the truth or appropriateness of the belief one may form, can give the subject a level of assurance.[347] Or, as Bergmann puts it, strong awareness makes it so that, from the subject's perspective, it would not be dumb luck that forming belief gets things correct. And so strong awareness can affect whether it is rational for the subject to form belief. But strong awareness is a kind of mental state of which we can ask whether the subject formed it rationally. For sometimes we irrationally conceive of things in the way we do. And since rationality requires awareness of reasons, we need to posit *other* states of strong awareness of reasons, if one rationally formed one's initial state of strong awareness. Thus if awareness is strong, foundationalism is in trouble.[348]

But if awareness of reasons is weak, then it does not involve conceiving of the reasons as in any way relevant to the truth or appropriateness of the belief one may form. And so it cannot give the subject some level of assurance. Put differently, it cannot help make it the case that, from the subject's perspective, it would not be dumb luck that forming belief gets things correct. Therefore, if the awareness is weak, there is no motivation for endorsing the Awareness

---

[346] Bergmann 2006, 11-13.
[347] On this point, see Fumerton (2010, section 3).
[348] Bergmann (2006, 14-19) stresses that the regress leads to skepticism. But it is clear that it also puts pressure on foundationalism.

Requirement, in the first place. Weak awareness appears to be quite impotent to affect rationality. So, if awareness is strong, foundationalism is in trouble; if weak, the Awareness Requirement is unmotivated.

There is a structural similarity between BonJour and Bergmann's dilemmas for the Awareness Requirement. On the one hand, consider awareness that is judgmental, or strong. These kinds of awareness appear to affect whether it is responsible to believe, or whether one has some level of assurance that forming belief would get things correct (and not by dumb luck). And so, such awareness appears to affect rationality. But at the same time, the question arises whether one was rational in forming states of awareness of these kinds. And so, it is unclear whether they can serve as foundational regress-stoppers. On the other hand, consider awareness that is nonjudgmental, or weak. Here the question of whether one was rational in forming the state of awareness does not arise. But at the same time, it is unclear whether such awareness affects whether it is responsible to believe, or whether one has some level of assurance that forming belief would get things correct (and not by dumb luck). And so, it is unclear whether such awareness affects rationality, and hence unclear whether there is any good motivation for requiring it.

These dilemmas split awareness in different ways. But they suggest that no matter how we slice up possible kinds of awareness, there is no kind of awareness of reasons that both plays the role in relation to the subject that motivates the Awareness Requirement in the first place *and* is such that the question of whether it was rationally formed does not arise. If so, then any kind of awareness either leaves the Awareness Requirement unmotivated or puts foundationalism at risk. We shall call this objection *the Dilemma*.

§10.4   How the Dilemma Threatens the Consciousness Requirement

Does the Dilemma threaten the Consciousness Requirement? Is there any reason to

suspect that no kind of awareness of reasons can both play the role in relation to the subject that

motivates the Consciousness Requirement and be such that the question of whether it was

rationally formed does not arise?

To begin, recall the role consciousness of reasons is supposed to play, and by virtue of

which the Consciousness Requirement has motivation. The fact that one is conscious of a reason

is supposed to affect what it is blameworthy (blameless) for one to do at the time. That is the role

consciousness of reasons plays on my picture.

So, is there reason to suspect that if consciousness of reasons affects what it is

blameworthy (blameless) for one to do at the time, then it is a mental state such that the question

of whether it was rationally formed arises? Yes. There is some initial plausibility to the thought

that a reason one does not so much as *believe* cannot affect what it is blameworthy (blameless)

for one to do at the time. For it is unclear how it is appropriate to blame others for failing to

respond to things that they do not believe. Indeed, several have given accounts of

blameworthiness on which one is blameworthy only if one has a particular kind of belief at the

time, e.g., the belief that they are doing wrong, or that they may harm someone, or etc.[349] It

seems that we have a firm grip on the idea that beliefs put us in contact with reasons in such a

way that our having this contact affects what would be blameworthy to do. But it seems we have

a tenuous grip, at best, on the idea that non-beliefs could do this. Thus, there is reason to suspect

that only conscious beliefs could, in the relevant way, affect what it is blameworthy (blameless)

---

[349]   See, e.g., Rosen (2004, 307) and Bergmann (2006, 90).

for one to do at the time. But, of course, a belief is a mental state such that the question of whether it was rationally formed arises.

So, the Dilemma has bite if framed in terms of the distinction between conscious beliefs and other states. And if the properties conscious beliefs have that explain why they affect blameworthiness are also properties of a state which make its formation evaluable for rationality, then the Dilemma has bite even if framed in other terms.

§10.5 Dissolving the Dilemma

While the Dilemma threatens the Consciousness Requirement, the accessibilist theory of which that requirement is a component has the resources to dissolve the Dilemma. Or so I shall argue. I shall begin by reminding us of the driving motivation for our accessibilist theory, which shall bring to light the question upon which hinges whether the Dilemma ultimately tells against the Consciousness Requirement.

Remember the driving motivation for our accessibilist theory, and so for its component Consciousness Requirement. On that motivation, rationality is the avoiding of presently available blameworthiness. But blameworthiness is presently available only if one is conscious of a set of reasons to which one can respond with an exercise of will, where an exercise of will is an event one experiences oneself as the source of amidst multiple alternatives also experienced as available to one at the time.

In light of this, the role consciousness of reasons plays, on my picture, is the role of presenting reasons *to which the will is responsive*. By the same token, the question of whether a mental state was rationally formed arises, in the relevant sense, only if its formation *can be an exercise of will*.

Therefore, whether the Dilemma for the Consciousness Requirement dissolves depends entirely on the following question: Is the will responsive to the objects of any kinds of conscious experiences or states such that the formation of those states cannot be an exercise of will? If so, then we not only have motivation to require these states, but, also, the question of whether these states were rationally formed does not arise. That is to say, the Dilemma dissolves if this question is answered affirmatively. At the same time, if the answer to this question is negative, then the Consciousness Requirement is in trouble.

With this question clear, we can point to a kind of conscious state to whose objects the will is responsive, even though the formation of that kind of state cannot be an exercise of will. The kind of state I have in mind is *conscious credences*.

To recall from Chapter 7, a credence in *p* is a feeling of *p* being true, which has a particular degree of strength, which may or may not be conscious at the time, and which helps explain certain dispositions, including the utility-risking dispositions one has concerning *p* given one's preferences. As I argued in Chapter 7, forming a credence can never be an exercise of will. So, the question of whether it is rational or irrational, at least in the relevant sense, cannot arise. But is the will responsive to reasons presented by credences? And is the will responsive to reasons presented by credences in cases in which the formation of a belief is an exercise of will?

The answer to each question is affirmative, at least when the relevant credence is a state of *consciousness*. When conscious, a credence presents a proposition in a particular light, as it is a conscious feeling of its being true, with a particular degree of strength. A conscious credence is, as it were, a consciousness of the probability of the proposition. And the will is responsive to probabilities so presented. Moreover, some of these cases include cases in which one's exercise of will is the forming of a belief.

For example, take the Biopsy Case from Chapter 9. Based on inductive evidence, you have a high credence that your skin spot is benign. But you have been told that the lab will call tomorrow evening and give you conclusive evidence one way or the other, and so you also have a high credence that the lab will call tomorrow evening. You are deliberating about whether to believe your skin spot is benign. As we think about this case, remember from Chapter 8 that to believe is to engage in the activity of organizing one's attention as if the proposition – that is, to attend only to the possibilities consistent with your skin spot's being benign as relevant to your reasoning.

Suppose that your high credence that your skin spot is benign, and your high credence that the lab will call tomorrow evening, are conscious. In being conscious of the high probability of your skin spot being benign, you may well have a nonconceptual, conscious sense of the availability of organizing your attention around the proposition that your skin spot is benign. You do not have a sense of this activity's unavailability, which you might have had if you were conscious of a low probability of your skin spot being benign. In any case, suppose that in this way, you entertain a content of a possible exercise of will, where this content is a way of beginning this activity. Moreover, in being conscious of the high probability that the lab will call tomorrow evening, you may well also be similarly conscious of the high probability that it is undesirable to organize one's attention around the proposition that the skin spot is benign, beforehand. At the same time, you may feel impatience, and just want to stop worrying about your health. If you realize the aforementioned content of the possible exercise of will, this exercise of will is the forming of the belief that your skin spot is benign. If you refrain, your reaffirming suspension is an exercise of will. And the important point is that, in this example, it is easy to suppose that *all* of the reasons to which your will responds are reasons presented to

you by conscious credences. It is easy to suppose that you exercise your will *on the basis* of only these probabilities of which you are conscious in this way.

Since conscious credences present reasons to which the will is responsive, they play the role consciousness of reasons needs to play, for my driving motivation for the Consciousness Requirement to have force. But since their formation cannot be an exercise of will, the question of whether they are rational, in the relevant sense, does not so much as arise. And hence foundationalism, which is meant to stop any regress of states evaluable for the relevant kind of rationality, is preserved. So, conscious credences show that the Dilemma for the Consciousness Requirement dissolves.

There may be lingering dissatisfaction, however. One might think that, on the above kind of reply to the Dilemma, the spirit of foundationalism has been sacrificed. For though not beliefs, credences are doxastic states, after all. And so when conscious credences are involved as described in belief formation, one might think that the belief is formed *by inference*. But one might have thought that the spirit of foundationalism includes the idea that rational *non-*inferential belief formation is possible.

In my own view, which I shall not take the space to defend, conscious credences are *not* the only kinds of conscious states in terms of which we can show that the Dilemma dissolves. But even if they were, the spirit of foundationalism is preserved. For remember that the driving motivation for foundationalism, as explained above. For a given kind of rationality, it is, intuitively, possible for it to be rational to form belief *without* needing to ride a wave of past irrationality, or even a wave of past events of forming states also evaluable for that kind of rationality. This motivation for foundationalism is preserved in my response to the Dilemma in terms of conscious credences. For they are not evaluable for the relevant kind of rationality. And to the complaint about the process being "inferential", belief formation always involves

transitions from antecedent mental states. What is central to foundationalism about rationality is that it is possible for these antecedent states to be not evaluable for the kind of rationality at issue.

A comparison can be made to Descartes' foundationalism. Descartes is a foundationalist about a kind of rationality – in particular, the rationality one has when one keeps one's belief within the boundaries of one's clear and distinct perceptions. But what kind of belief does this kind of foundationalism concern? Well, it is the kind of belief Descartes' skeptical reflections were meant to dislodge. And plausibly, these beliefs are *not* credences but rather outright doxastic states of some kind.

Here is why I say this. Observe Descartes:

> 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 never get out of the habit of confidently assenting to these opinions, so long as I suppose them to be what in fact they are, namely highly probable opinions – opinions which, despite the fact that they are in a sense doubtful, as has just been shown, it is still much more reasonable to believe than to deny. In view of this, I think it will be a good plan to turn my will in completely the opposite direction and deceive myself, by pretending for a time that these former opinions are utterly false and imaginary.[350]

This paragraph evinces the fact that Descartes's skeptical reflections have not eliminated his fairly high credence in, e.g., propositions about the external world. He admits frankly that his former opinions are "highly probable" and "much more reasonable to believe than to deny". But notice the following: he does *not* take it on himself to purge the credences herein reported. Rather, he points out that the high probability of propositions about the external world make it difficult for him to cease believing them, especially given his long habit of doing so. And then he says that, as a strategy to purge all uncertain beliefs, he will *pretend* his former opinions are

---

[350]   Descartes 1986 [1641], 15.

"utterly false and imaginary". This pretense, in effect, comes to him focusing intently on those possibilities in which feature the negations of the propositions he has a fairly high credence in. His acknowledgment of pretense underscores that he has fairly high credence in these propositions. Simultaneously, he is *unconcerned* that it is pretense, even promising the pretense is but "for a time". This is not what someone would say who is intent on purging himself of his fairly high credences. That kind of philosopher would aim to come to the point where no pretense is necessary, and, instead, nothing appears more probable than anything else. Descartes has no such aim.

Thus, what Descartes is intent on purging himself of is not his fairly high credences but instead uncertain *outright* belief. It is of belief so understood which his foundationalism about rationality concerns. The same is true of my foundationalism about rationality. But where, in his obsession with certainty, Descartes found his fairly high credences to be a distraction, we can allow for cases in which conscious credences present to us foundations for rational belief formation. In doing this, we allow for cases in which forming belief is rational in a sense of 'rational' in which none of the other states in play at the time are evaluable for rationality. The spirit of foundationalism about rationality is that *this* kind of starting point for rational belief formation is available. For it is this kind of starting point that allows for a kind of non-question-begging critique of the rationality of one's other beliefs.

We have seen above how the Dilemma for the Consciousness Requirement dissolves. But to secure the point, let us return to the instances of the Dilemma offered by BonJour and Bergmann.

According to BonJour's dilemma, awareness is either judgmental or nonjudgmental. Nonjudgmental awareness is supposed to be impotent to affect what it is responsible to believe while. On the other hand, judgmental awareness is a state whose formation may or may not be

responsible, and so is evaluable for the relevant kind of rationality. But even if conscious credences are "judgmental", we have seen that they are not evaluable for the relevant kind of rationality – viz., blamelessness. So, these states present reasons which affect that kind of rationality without being states evaluable for it.

According to Bergmann's dilemma, awareness is either strong (i.e., involves conceiving of the reason as relevant to the appropriateness or truth of the available belief) or weak (i.e., not strong). Weak awareness is supposed to be impotent to give the subject any level of assurance, i.e., make it so that, from the subject's perspective, it would not be dumb luck that forming belief gets things correct. But even if some conscious credences constitute strong awareness of reasons, once again, we have seen that they are not evaluable for the relevant kind of rationality. The strong horn of Bergmann's dilemma, then, does not in fact beget regress.

§10.6   Conclusion

I began by explaining how the Consciousness Requirement is a built-in component of the full-fledged accessibilist theory for which this dissertation has provided a deontological explanation. But the Consciousness Requirement is a special case of the Awareness Requirement, and several have argued that the Awareness Requirement is either unmotivated or in conflict with foundationalism. We saw that this kind of dilemma threatens the Consciousness Requirement, in that there is some initial plausibility to the idea that a blameworthy alternative is available to one only if one already has certain beliefs at the time. But when we returned again to our motivation for the Consciousness Requirement, we saw that the Dilemma for the Consciousness Requirement dissolves. For the will is responsive to more than just the objects of belief. It is also responsive to the reasons presented by conscious credences. At the same time, conscious credences are not evaluable for rationality, because their formation cannot be an

exercise of will. So, this kind of consciousness of reasons plays the role consciousness of reasons needs to play, to motivate the Consciousness Requirement. Yet they are not evaluable for rationality, and so foundationalism is preserved.

In conclusion, I shall make a final point. My dissertation includes no substantive argument that deontologism is true. However, it has developed an overall deontological picture that may allow us to check and see what deontologism has going for it. On this picture, the question of blameworthiness arises only when forming a belief is an exercise of will. And we have seen why many formings of belief are indeed exercises of will, in light of the presence of practical epistemic reasons, and given a restriction to outright belief. With this picture in hand, we can ask two questions. Are there significant uses of 'rational' and 'irrational' that align with 'blameless' and 'blameworthy' as those terms are understood in this overall picture? And if so, do these uses match the intuitive motivations originally on offer for deontologism? We can look and see, by checking whether these things are so, whether deontologism enjoys a level of vindication.

As it turns out, to return to the intuitive motivations originally on offer for deontologism is to bring to the fore the very kinds of cases in which it seems natural to use 'rational' and 'irrational' in the sense at issue. Consider, for example, W.K. Clifford's famous example of the Shipowner, which is meant to motivate deontologism.[351] The Shipowner has some doubts whether his aging vessel is still sea-worthy. But he ignores these doubts, and reminds himself how much inductive evidence he has of the old ship's sea-worthiness. He comes to believe the ship is safe, and so omits refitting the hull. This is a case in which he has strong practical epistemic reason to suspend, and in which his forming belief might well be an exercise of will.

---

[351]   Clifford 1877, 289-290.

With a relatively high credence that the ship is safe, he experiences as available to him the initial phase of the activity of organizing his attention as if the ship is safe. But, given his strong practical epistemic reasons to suspend, he also experiences refraining from this activity as available to him. Thinking of the case in this way, his forming belief is an exercise of will. Simultaneously, it is a forming of a belief that we may call 'irrational!', and in a sense in which we express blame.

So, it appears that our overall deontological picture has latched onto a *kind* of epistemic rationality. And that is as much as I can hope for, in this dissertation. Many questions are left open for future research. Chief among them, perhaps, is the question of how this kind of epistemic rationality relates to (various kinds of) knowledge.

Bibliography

Alston, William. (1989) *Epistemic Justification: Essays in the Theory of Knowledge* (London: Cornell University Press)

Anscombe, Elizabeth. (1969 [1957]) *Intention* (New York: Cornell University Press)

Armstrong, D.M. (1973) *Belief, Truth, and Knowledge* (Cambridge: Cambridge University Press)

Armstrong, Katherine M. (2011) "Covert Spatial Attention and Saccade Planning", in Mole, Smithies, and Wu (eds) *Attention: Philosophical and Psychological Essays* (Oxford: Oxford University Press)

Arvidson, Sven. (1996) "Towards a Phenomenology of Attention", *Human Studies*, 19 (1): 71-84.

Audi, Paul. (2012) "Grounding: Toward a Theory of the In-Virtue-Of Relation" *Journal of Philosophy*, 109 (12): 685-711.

Audi, Robert. (1994) "Dispositional Beliefs and Dispositions to Believe", *Noûs* 28 (4): 419-34.

__ (1998) *Epistemology: A Contemporary Introduction to the Theory of Knowledge* (New York: Routledge)

Bailey, Andrew. (2015) "The Priority Principle", *Journal of the American Philosophical Association*, 1 (1): 163-174.

Baker, L. (2007) "Persons and other things", *Journal of Consciousness Studies* 14: 17–36.

Balaguer, Mark. (2004) "A Coherent, Naturalistic, and Plausible Formulation of Libertarian Free Will", *Noûs* 38 (3): 379–406.

Ball, Brian and Blome-Tillmann, Michael. (2013) "Indexical Reliabilism and the New Evil Demon", *Erkenntnis* 78 (6): 1317-1336.

Bargh, John A., Chen, Mark, and Burrows, Lara. (1996) "Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action", *Journal of Personality and Social Psychology* 71 (2): 230-244.

Bayne, Tim, and Montague, Michelle (eds.). (2011) *Cognitive Phenomenology* (Oxford: Oxford University Press)

Bennett, Jonathan. (2002) "What Events Are", *The Blackwell Guide to Metaphysics*, 43-65.

Bergmann, Michael. (2006) *Justification without Awareness* (Oxford: Oxford University Press)

Block, Ned. (1995) "On a Confusion about a Function of Consciousness", *Brain and Behavioral Sciences* 18 (2): 227-247.

BonJour, Laurence. (1985) *The Structure of Empirical Knowledge* (Cambridge, Massachusetts: Harvard University Press)

Boethius. *Liber de Persona et Duabus Naturis*

Boyle, Matthew. (2011) "'Making Up Your Mind' and the Activity of Reason", *Philosophers' Imprint*, Volume 11, No. 17: 1-24.

Bratman, Michael. (1987) *Intention, Plans, and Practical Reason* (Harvard University Press)

__ (1992) "Practical Reasoning and Acceptance in a Context", *Mind*, 101: 1-15.

Broome, John. (2013) *Rationality Through Reasoning* (John Wiley & Sons)

Broughton, R., Billings, R., et al. (1994) "Homicidal Somnambulism: A Case Report", *Sleep 17* (3): 253-264.

Brown, Jessica. (2007) "Externalism in Mind and Epistemology", in Goldberg, Sanford C. (ed) *Internalism and Externalism in Semantics and Epistemology* (Oxford: Oxford University Press)

Carroll, Lewis. (1895) "What the Tortoise Said to Achilles", *Mind* 4 (14): 278-280.

Caruso, G. D. (2012) *Free Will and Consciousness: A Determinist Account of the Illusion of Free Will* (New York: Lexington Books)

__ (2015) "If Consciousness is Necessary for Moral Responsibility, then People are Less Responsible than We Think", *Journal of Consciousness Studies*, 22 (7-8): 49-60.

Casati, Roberto, and Varzi, Achille. (2014) "Events", in Zalta, Edward N. (ed.) *The Stanford Encyclopedia of Philosophy* URL = <http://plato.stanford.edu/archives/win2015/entries/events/>.

Chalmers, David. (1996) *The Conscious Mind* (Oxford: Oxford University Press)

Clifford, W.K. (1877) "The Ethics of Belief, repr.", *The Ethics of Belief Debate*.

Cohen, Jonathan L. (1989) "Belief and Acceptance", *Mind*, Vol. 98, Issue 391: 367-389.

Cohen, Stewart. (1984) "Justification and Truth", *Philosophical Studies* 46 (3): 279-295.

Conee, Earl, and Feldman, Richard. (2001) "Internalism Defended", *American Philosophical Quarterly*, 38 (1): 1-18.

Crane, Tim. (1991) "All the Difference in the World", reprinted in: Pessin, Goldberg, and Putnam (eds.) (1996) *The Twin Earth Chronicles*, 284-304 (New York: M.E. Sharpe)

Dallmann, Justin M. (2014) "A Normatively Adequate Credal Reductivism", *Synthese*, 19 (10): 2301-2313.

Davidson, Donald. (1963) "Actions, Reasons, and Causes", *The Journal of Philosophy* 60: 685-700.

Descartes, Rene. (1986 [1641]) *Meditations on First Philosophy*, translated by John Cottingham (Cambridge University Press)

Dillon, R. (2007) "Arrogance, self-respect and personhood", *Journal of Consciousness Studies* 14, 101–126.

Doring, Sabine A. (2007) "Seeing What to Do: Affective Perception and Rational Motivation", *Dialectica*, 61, 3, pp. 363-94.

Eriksson, Lina and Hajek, Alan. (2007) "What Are Degrees of Belief?", *Studia Logica: An International Journal for Symbolic Logic*, 86 (2): 183-213.

Fales, Evan. (2014) "Turtle Epistemology", *Philosophical Studies* 169 (2): 339-354.

Fantl, Jeremy and McGrath, Matthew. (2010) *Knowledge in an Uncertain World* (Oxford: Oxford University Press)

Farkas, Katalin. (2003) "Does Twin Earth Rest on a Mistake?", *Croatian Journal of Philosophy* 3 (2): 155-169.

Feldman, Richard. (2001) "Voluntary Belief and Epistemic Evaluation", in Steup, Matthias (ed.) *Knowledge, Truth, and Duty* (Oxford: Oxford University Press)

Ferrero, Luca. (2002) *Making Up One's Self: Agency, Commitments and Identity* (Doctoral Dissertation, Harvard University, Cambridge, Massachusetts)

__ (2009) "Conditional Intentions", *Noûs* 43 (4): 700-741.

Fischer, John and Ravizza, Mark. (1998) *Responsibility and Control* (New York: Cambridge University Press)

__ (1991) "Responsibility and inevitability", *Ethics* 101, 258–278.

Foley, Richard. (2009) "Beliefs, Degrees of Belief, and the Lockean Thesis", in Huber and Schmidt-Petri (eds.) *Degrees of Belief,* 37-47 (Springer: Synthese Library)

Foster, J. (2002) *The immaterial self: a defence of the Cartesian dualist conception of the mind* (London: Routledge)

Frankish, Keith. (2004) *Mind and Supermind* (Cambridge: Cambridge University Press)

__ (2007) "Deciding to Believe Again", *Mind* 116 (463): 523-547.

__ (2009) "Partial Belief and Flat-Out Belief", in Huber and Schmidt-Petri (eds.) *Degrees of Belief,* 75-93 (Springer: Synthese Library)

Frankfurt, H. G. (1969) "Alternate Possibilities and Moral Responsibility", *The Journal of Philosophy*, *66* (23): 829-839.

__ (1971) "Freedom of the Will and the Concept of a Person" *Journal of Philosophy*, 68 (1): 5–20.

Fumerton, Richard. (1995) *Metaepistemology and Skepticism* (Maryland: Rowman & Littlefield Publishers, Inc.)

__ (2010) "Foundationalist Theories of Epistemic Justification", in Zalta, Edward N. (ed.) *The Stanford Encyclopedia of Philosophy* URL = <http://plato.stanford.edu/archives/sum2010/entries/justep-foundational/>.

Ginet, Carl. (1975) *Knowledge, Perception, and Memory* (Dordrecht: D. Reidel Publishing Company)

__ (1990) *On Action* (New York: Cambridge University Press)

Goldman, Alvin. (1976) "Discrimination and Perceptual Knowledge", *The Journal of Philosophy*, 73 (20): 771-791.

Greco, Daniel. (2014) "Could KK Be OK?", *The Journal of Philosophy*, 111 (4): 169-197

Greco, John. (2010) *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity* (New York: Cambridge University Press)

Hansson, Sven Ove, and Grüne-Yanoff, Till. (2012) "Preferences", in Zalta, Edward (ed.) *The Stanford Encyclopedia of Philosophy*, URL = <http://plato.stanford.edu/archives/win2012/entries/preferences/>.

Hawthorne, John. (2004) *Knowledge and Lotteries* (Oxford: Oxford University Press)

Hieronymi, Pamela. (2005) "The Wrong Kind of Reason", *The Journal of Philosophy*, 102 (9): 437-457.

__ (2006) "Controlling Attitudes", *Pacific Philosophical Quarterly* 87 (1): 45-74.

__ (2013) "The Use of Reasons in Thought (and the Use of Earmarks in Arguments)", *Ethics*, 124 (1): 114-127.

Hopp, Walter. (2011) *Perception and Knowledge* (New York: Cambridge University Press)

Horgan, Terry. (2011) "The Phenomenology of Agency and Freedom: Lessons From Introspection and Lessons From Its Limits", *Humana Mente* 15: 77-97.

Horgan, Terry, and Tienson, J. (2002) "The Intentionality of Phenomenology and the Phenomenology of Intentionality" in Chalmers, David (ed.) *Philosophy of Mind: Classical and Contemporary Readings,* 520-533 (Oxford University Press)

Hubbs, Graham. (2013) "How Reasons Bear on Intentions", *Ethics*, *124* (1): 84-100.

James, William. (1950) *The Principles of Psychology*, *Vol. Two* (New York: Dover Publications, Inc.)

Johnston, M. (2001) "The Authority of Affect", *Philosophy and Phenomenological Research*, 63 (1): 181-214.

Kane, Robert. (1985) *Free Will and Values*: *Adaptive Mechanisms and Strategies of Prey and Predators* (SUNY Press)

Kant, I. (2002 [1785]) *Groundwork for the Metaphysics of Morals* (translation by Allen Wood) (London: Yale University Press)

Kavka, Gregory S. (1983) "The Toxin Puzzle", *Analysis*, 43 (1): 33-36.

Keller, I., Heckhausen, H. (1990) "Readiness potentials preceding spontaneous motor acts: voluntary vs. involuntary control", *Electroencephalogr Clin Neurophysiol.* 76 (4): 351-361.

Kelly, Thomas. (2002) "The Rationality of Belief and Some Other Propositional Attitudes", *Philosophical Studies*, *110* (2): 163-196.

Kenny, Anthony. (1976) *Will, Freedom, and Power* (Blackwell Publishers)

Kim, Jaegwon. (1966) "On the Psycho-Physical Identity Theory", *American Philosophical Quarterly*, 3 (3): 227-235.

__ (2002) *Supervenience* (Vermont: Ashgate Publishing Company)

King, M.; Carruthers, P. (2012) "Moral Responsibility and Consciousness", *Journal of Moral Philosophy* 9:200-228.

Kornhuber, H.H., Deecke, L. (1964) "Hirnpotentialänderungen beim Menschen vor und nach Willkürbewegungen, dargestellt mit Magnetbandspeicherung und Rückwärtsanalyse", *Pflügers Arch Eur J Physiologie* 281: 52.

Levin, Janet. (2013) "Functionalism", in Zalta, Edward (ed) *The Stanford Encyclopedia of Philosophy*, URL = <http://plato.stanford.edu/archives/fall2013/entries/functionalism/>.

Levy, Neil. (2014) *Consciousness and Moral Responsibility* (Oxford: Oxford University Press)

__ (2015) "Defending the Consciousness Thesis: A Response to Robichaud, Sripada and Caruso", *Journal of Consciousness Studies*, 22 (7-8): 61-76.

Lewis, David. (1986) *On the Plurality of Worlds* (Oxford: Blackwell)

Locke, J. (2009 [1690]) *Essay Concerning Human Understanding*, in Ariew (ed) *Modern Philosophy: An Anthology of Primary Sources* (Cambridge: Hackett Publishing Company)

Lyons, Jack. (2009) *Perception and Basic Beliefs* (Oxford: Oxford University Press)

McCain, Kevin. (2015) "A New Evil Demon? No Problem for Moderate Internalists", *Acta Analytica*, 30 (1): 97-105.

Moon, Andrew. (2012) "Three Forms of Internalism and the New Evil Demon Problem", *Episteme*, 9 (4): 345-360.

__ (2015) "The New Evil Demon, a Frankfurt-style Counterfactual Intervener, and a Subject's Perspective Objection: Reply to McCain", *Acta Analytica*, *30* (1): 107-116.

Montmarquet, James A. (1993) *Epistemic Virtue and Doxastic Responsibility*. (Lanham New Jersey)

Mele, Alfred and Moser, Paul. (1994) "Intentional Action", *Noûs* 28 (1): 39-68.

Mole, Christopher. (2011) "The Metaphysics of Attention", in Mole, Smithies, and Wu (eds) *Attention: Philosophical and Psychological Essays* (Oxford: Oxford University Press)

Moran, Richard, and Stone, Martin. (2011) "Anscombe on Expression of Intention: An Exegesis", in (eds. Hornsby and Stoutland) *Essays on Anscombe's Intention* (Harvard University Press)

Mourad, Ronney. (2008) "Choosing to Believe", *International Journal for the Philosophy of Religion*, Vol. 63, No. 1/3, *Ethics of Belief: Essays in Tribute to D. Z. Philips*: 55-69.

Murray, D., and Nahmias, E. (2014) "Explaining Away Incompatibilist Intuitions", *Philosophy and Phenomenological Research* 88, 434–467.

Nagel, T. (1974) "What is It Like to Be a Bat?", *Philosophical Review* 83: 435-50.

__ (1993) "Moral Luck", in Statman, Daniel (ed.) *Moral Luck* (SUNY Press)

Nichols, Shaun. (2012) "The Indeterminist Intuition: Source and Status", *The Monist*: 290-307.

Nottelmann, Nikolaj. (2007) *Blameworthy Belief: A Study in Epistemic Deontologism* (Springer Science & Business Media, Vol. 338)

O'Connor, Timothy. (2000) *Persons and Causes: The Metaphysics of Free Will* (Oxford University Press)

Olson, E.T. (2007) *What Are We?* (Oxford: Oxford University Press)

O'Shaughnessy, Brian. (1980) *The Will: A Dual Aspect Theory, Vol. 1-2* (Cambridge University Press)

Parfit, Derek. (2001) "Rationality and Reasons" *Exploring Practical Philosophy: From Action to Values*, 17-39.

Pillar, Christian. (2001) "Normative Practical Reasoning" in *Aristotelian Society Supplementary Volume* 75 (1): 195-216.

Pitt, David. (2009) "Intentional Psychologism", *Philosophical Studies*, *146* (1): 117-138.

Plantinga, Alvin. (1993) *Warrant: The Current Debate* (New York: Oxford University Press)

Podgorski, Abelard. (2016a) "A Reply to the Synchronist", *Mind* 125 (499): 859-871.

__ (2016b) "Dynamic Permissivism", *Philosophical Studies* 173 (7): 1923-1939.

Pryor, Jim. (2001) "Highlights of Recent Epistemology", *British Journal for the Philosophy of Science* 52 (1): 95-124.

Putnam, Hilary. (1975) "The Meaning of 'Meaning'," reprinted in: Pessin, Goldberg, and Putnam (eds) (1996) *The Twin Earth Chronicles*, 3-52 (New York: M.E. Sharpe)

Rabinowicz, Wlodek and Ronnow-Rasmussen, Toni. (2004) "The Strike of the Demon: On Fitting Pro-Attitudes and Value" *Ethics*, 114 (3): 391-423.

Ramsey, F.P. (1926) "Truth and Probability," Ch. VII in Braithwaite (ed) *The Foundations of Mathematics and Other Logical Essays* (New York: Harcourt, Brace and Company)

Rosen, Gideon. (2004) "Skepticism about Moral Responsibility", *Philosophical Perspectives*, 18 (1): 295-313.

Ross, Jacob, and Mark Schroeder. (2014) "Belief, Credence, and Pragmatic Encroachment", *Philosophy and Phenomenological Research*, 88 (2): 259-288.

Ruff, Christian C. (2011) "A Systems-Neuroscience View of Attention", in Mole, Smithies, and Wu (eds) *Attention: Philosophical and Psychological Essays* (Oxford University Press)

Ryan, Sharon. (2003) "Doxastic Compatibilism and the Ethics of Belief", *Philosophical Studies*, 114 (1): 47-79.

Ryle, Gilbert. (2001 [1949]) *The Concept of Mind* (Chicago: The University of Chicago Press)

Schantz, Richard. (ed.) (2004) "Introduction", in *The Externalist Challenge* Vol 2, (Walter de Gruyter)

Schroeder, Mark. (2007) *Slaves of the Passions* (New York: Oxford University Press)

__ (2008) "Having Reasons", *Philosophical Studies* 139 (1): 57-71.

__ (2010) "Value and the Right Kind of Reason", *Oxford Studies in Metaethics* 5:25-55.

__ (2011) "*Ought*, Agents, and Actions", *Philosophical Review* 120 (1): 1-41.

__ (2012a) "The Ubiquity of State-Given Reasons", *Ethics* 122 (3): 457-488.

__ (2012b) "Stakes, Withholding, and Pragmatic Encroachment on Knowledge", *Philosophical Studies*, 160 (2): 265-285.

__ (2013) "State-Given Reasons: Prevalent, If Not Ubiquitous", *Ethics*, 124 (1): 128-140.

Searle, John. (1984) *Minds, Brains, and Science* (Harvard University Press)

__ (2001) *Rationality In Action* (Cambridge: MIT Press)

__ (2004) "Minds, Brains, and Programs", in Shieber, Stuart (ed.), 201-224 *The Turing Test* (MIT Press)

Setiya, Kieran. (2007) *Reasons Without Rationalism* (New Jersey: Princeton University Press)

Shah, Nishi, and Silverstein, Matthew. (2013) "Reasoning in Stages", *Ethics*, 124 (1): 101-113.

Shah, Nishi. (2003) "How Truth Governs Belief", *The Philosophical Review*, 112 (4): 447-482.

__ (2006) "A New Argument for Evidentialism", *The Philosophical Quarterly*, *56* (225): 481-498.

Sher, George. (2009) *Who Knew? Responsibility Without Awareness* (Oxford: Oxford University Press)

Smart, J. J. C. (1961) "Free-will, Praise and Blame", *Mind,* 70 (279): 291-306.

Smith, Michael. (1987) "The Humean Theory of Motivation", *Mind,* 96 (381): 36-61.

Srinivasan, Amia. (2015) "Are We Luminous?", *Philosophy and Phenomenological Research*, 90 (2): 294-319.

Sripada, C. (2015) "Acting from the Gut: Responsibility without Awareness", *Journal of Consciousness Studies*, 22 (7-8): 37-48.

Stalnaker, Robert. (1984) *Inquiry* (MIT Press)

Steup, Matthias. (2005) "Epistemology", in Zalta, Edward N. (ed.), *The Stanford Encyclopedia of Philosophy* URL = <http://plato.stanford.edu/archives/spr2014/entries/epistemology/>.

Stone, J. (1988) "Parfit and the Buddha: Why There Are No People", *Philosophy and Phenomenological Research* 48, 519.

Strawson, P. F., 1962, "Freedom and Resentment", *Proceedings of the British Academy*, 48: 1–25.

Thompson, Michael. (2008) *Life and Action: Elementary Structures of Practice and Practical Thought* (Harvard University Press)

Turri, John. (2010) "On the Relationship between Propositional and Doxastic Justification", *Philosophy and Phenomenological Research*, 80 (2): 312-326.

Twain, Mark. (1885) *The Adventures of Huckleberry Finn* (Charles L. Webster and Company)

Van Cleve, James. (1990) "Supervenience and Closure", *Philosophical Studies*, 58: 225-238

Vargas, Manuel. (2005) "The Trouble with Tracing", *Midwest Studies in Philosophy*, Vol. 29, No. 1: 269-291.

Velleman, J. David. (1992) "What Happens When Someone Acts?", *Mind*, 101 (403): 461-481.

Wallace, R. Jay. (1994) *Responsibility and the Moral Sentiments* (Harvard University Press)

Watzl, Sebastian. (2011) "Attention as Structuring of the Stream of Consciousness", in Mole, Smithies, and Wu (eds) *Attention: Philosophical and Psychological Essays* (Oxford: Oxford University Press)

Weatherson, Brian. (2005) "Can We Do Without Pragmatic Encroachment?", *Philosophical Perspectives*, 19: 417-443.

Wedgwood, Ralph. (2001) "Conceptual Role Semantics for Moral Terms", *Philosophical Review* 110: 1-30.

__ (2002a) "Internalism Explained", *Philosophy and Phenomenological Research*, 65 (2): 349-369.

__ (2002b) "The Aim of Belief", *Philosophical Perspectives* 36 (16): 267-297.

__ (2006) "The Normative Force of Reasoning", *Nous*, 40 (4): 660-686

__ (2007) *The Nature of Normativity* (New York: Oxford University Press)

__ (2009) "Intrinsic Values and Reasons for Action", in Sosa, Ernest and Villanueva, Enrique (eds.) *Metaethics* (Wiley Periodicals, Inc.), 342-363.

__ (2012) "Outright Belief", *Dialectica* 66 (3): 309-329.

__ (2013a) "Rational 'Ought' Implies 'Can'", *Philosophical Perspectives*, 23: 70-92.

__ (2013b) "Doxastic Correctness", *Proceedings of the Aristotelian Society*, Supplementary Vol. 87: 38-54.

Willard, Dallas. (1967) "A Crucial Error in Epistemology", *Mind*, 76: 513-523.

Williams, Bernard. (1973) "Deciding to Believe", in *Problems of the Self: Philosophical Papers 1956-1972* (Cambridge: Cambridge University Press)

__ (1979) "Internal and External Reasons", in Harrison, Ross (ed.) *Rational Action* (Cambridge University Press), 101-113.

__ (1995) "Internal Reasons and the Obscurity of Blame", in *Making Sense of Humanity* (Cambridge: Cambridge University Press)

Williamson, Timothy. (2000) *Knowledge and Its Limits* (Oxford: Oxford University Press)

__ (2013) "Response to Cohen, Comesaña, Goodman, Nagel, and Weatherson on Gettier Cases in Epistemic Logic", *Inquiry* 56 (1): 77-96.

Worsnip, Alexander. (2015) "Rationality's Demands on Belief", Ph.D Dissertation, Yale University.

Wu, Wayne. (2011) "Attention as Selection for Action", in Mole, Smithies, and Wu (eds) *Attention: Philosophical and Psychological Essays* (Oxford: Oxford University Press)

Zhu, Jing. (2003) "The Conative Mind: Volition and Action", PhD Dissertation, University of Waterloo.