UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Fundamentalist Contextualist Compatibilism: A Response to the Consequence Argument

Permalink

https://escholarship.org/uc/item/0tg2c5nq

Author

Pendergraft, Garrett

Publication Date

2010

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Fundamentalist Contextualist Compatibilism: A Response to the Consequence Argument

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Philosophy

by

Garrett Heath Pendergraft

December 2010

Dissertation Committee:
Dr. John Martin Fischer, Chairperson
Dr. Peter J. Graham
Dr. John R. Perry

Copyright by
Garrett Heath Pendergraft
2010

The Dissertation of Garrett Heath Pendergraft is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgments

The first (and last) acknowledgment belongs to my Lord and Savior, Jesus Christ: the wisest man, and the greatest philosopher, who ever lived.

My journey through graduate school included several twists and turns, and I was blessed with excellent teaching and guidance at every stop along the way. At Biola University I benefited from the influence of Garry DeWeese, Doug Geivett, and David Hunt, who (as a part of the MA Phil program) set me on the path that led to today. At the University of Missouri-Columbia, I benefited greatly from the tutelage of André Ariew, Jon Kvanvig, Andrew Melnyk, Matt McGrath, Peter Vallentyne (without whom I probably wouldn't have ended up at UCR), and Paul Weirich. I also benefited from conversations and fellowship with other graduate students—including Kenny Boyce, Eric Heidenreich, Justin McBrayer, Kevin McCain, Andrew Moon, Ted Poston, Brandon Schmidly (my carpool partner), and Patrick Todd.

At the University of California-Riverside, I was fortunate to be in an environment that provided not only an excellent education in philosophy, but also the support and kindness that every academic department should strive for. From UCR, thanks are due to several of the many outstanding faculty there: Peter Graham, Robin Jeshion, Pierre Keller (my faculty mentor), Coleen Macnamara, Michael Nelson, John Perry, Erich Reck (who ably supervised my Proposition), Eric Schwitzgebel, Gary Watson, and Howie Wettstein. Thanks also to the office staff, and in particular Jayne Gales. At UCR I benefited from being a part of a wonderful community of grad students, including Justin Coates, Joe Cressotti (my office-mate), Chris Franklin, Ben Mitchell-Yellin, Luis Montes, Scott Sevier, Philip Swenson, Patrick Todd (again!), and Neal Tognazzini. And I especially must single out my advisor, John Fischer. In addition to being a first-rate philosopher, John was more generous, patient, and supportive than I had ever imagined an advisor could be. He has been a

philosophical mentor to me, and a good friend, and I owe my success in the profession to him more than anyone else.

I owe final thanks to my family, both immediate and extended. They saw me through, and endured the multifarious indirect effects of, my long and tortuous graduate school journey. They have been unflagging in their support and profligate in their encouragement. My wife, Amy, has been my partner and confidante—my best friend—throughout the entire process, and it is literally true that I could not have done this without her unfailing love and encouragement. My children, Jenica Sage and Curran Landry, also contributed mightily to this project—in ways they (and even I) don't fully understand. And Sage, to answer your question: Yes, it's done!

Chapter 3, in part, is a reprint of material that will appear in *Philosophical Studies,* forthcoming (published online: July 13, 2010). I acknowledge and thank the publisher for permission to reuse this material.

Dedication

To my father: Gary Brian Pendergraft (1951–1996).

ABSTRACT OF THE DISSERTATION

Fundamentalist Contextualist Compatibilism:
A Response to the Consequence Argument

by

Garrett Heath Pendergraft

Doctor of Philosophy, Graduate Program in Philosophy
University of California, Riverside, December 2010
Dr. John Martin Fischer, Chairperson

In my dissertation I offer what I take to be a novel and compelling response to the *consequence argument*: the argument that if causal determinism is true, then the past history of the world and the laws of nature together determine everything that will happen in the future—including my actions and in fact every action ever done by anyone. I begin by noting and emphasizing a parallel between the consequence argument and the skeptical argument, which leads us to ask whether a response to the latter can be modified and applied to the former. In preparation for that undertaking, we examine two influential responses to the consequence argument—*backtracking compatibilism* and *local miracle compatibilism*—both of which claim that if we were to do otherwise (and if determinism is true), then a certain counterfactual conditional would be true. Although I don't fully endorse either of these responses, I do explain how they point us in the right direction.

I then turn to the skeptical argument, and in particular the *contextualist* response to the skeptical argument. Although I don't fully endorse contextualism either, I do emphasize a virtue of the view, namely that it explains how the skeptical argument can seem so compelling even though, in ordinary circumstances, its conclusion strikes us as wildly implausible.

Finally, I offer my response to the consequence argument. I begin by adopting and extending a philosophical methodology labeled "southern fundamentalism." The first move in my response is to argue that we should endorse an "austere" conception of acting freely

according to which it does not require being able to do otherwise than we actually do, as an extension of the actual past (consistent with the laws of nature). I then provide a contextualist explanation of how we can be led (astray) by the consequence argument into thinking that this condition is required for acting freely when in fact it is not.

Thus I hope to have provided not only a new and compelling response to the consequence argument, but also a foray into some woefully under-explored territory: the intersection of agency theory and epistemology.

Table of Contents

CHAPTER 7: THE EPISTEMOLOGICAL TREATMENT

List of figures

• Chapter 1 •

A tale of two arguments

1.1 Introduction

Consider a scenario that occurs all too often: I raise my coffee cup to take a drink. This action of mine intuitively seems to have been a *free action*. I raised my cup, but I could have refrained; whether I took that drink at that time was up to me. That particular free action was perhaps inconsequential, but I have performed other free actions that were much more consequential. And in any case all of my free actions are important insofar as they contribute to making me the unique person that I am. If I were to discover that a significant portion of these actions weren't free after all, then it seems that I wouldn't be the person I thought I was. And if I were to discover that *none* of them were free, then it seems that I might not be a person at all. Unfortunately, there's a powerful argument for precisely this conclusion—that none of my actions are free—and moreover that none of anybody's actions are ever free. This argument is called the *consequence argument*, and it is the argument that I will be focusing on in this dissertation.[1] In this chapter I will introduce it briefly and then provide a semi-formal presentation that we will be working with in what follows.

1.2 The consequence argument

The consequence argument was introduced by David Wiggins (2003)[2] and Carl Ginet (1990),[3] but languished in relative obscurity until Peter van Inwagen's (1983)[4] presentation

[1] It would be more precise to say that the label "the consequence argument" refers to a *family* of arguments—one member of which we will be focusing on in what follows.

[2] Wiggins (2003) contains material dating back to 1965.

brought it to prominence in the philosophical literature. This argument comes in different versions, but they all purport to capture the same basic idea, which is this: If determinism is true, then my actions are nothing more than the consequences of past events (which were governed by the laws of nature). And if my actions are nothing more than the consequences of past events, then I had no control over them. Hence, my actions are not free (and neither are anybody else's). In other words, if determinism is true then free will does not exist; freedom is incompatible with determinism.

What we learn from the consequence argument, then, is that causal determinism threatens—and perhaps even precludes—free will. But what exactly is causal determinism? This is a surprisingly difficult question to answer,[5] but I will try to say enough to evince a basic understanding of the concept. (A *deep* understanding of determinism would require a dissertation-length project in itself, so we'll have to be content with something less than that.[6]) According to what I take to be a fairly standard and relatively uncontroversial definition of causal determinism, an event is causally determined if it is entailed by the laws of nature and the history of the world prior to that event:

> Causal determinism is the thesis that, for any given time, a complete statement of the facts about that time, together with a complete statement of the laws of nature, entails every truth as to what happens after that time. (Fischer 1995, 9)

(For the sake of brevity, I will typically drop the "causal" and simply refer to the above thesis as the thesis of determinism.) According to this definition, the force with which past events

---

[3] Ginet (1990, 94n4) describes the genealogy of his formulation of the argument as follows: "The argument I will present is a descendant of one I presented in Ginet (1966), which had other, somewhat closer descendants in Ginet (1980) and Ginet (1983)."

[4] Van Inwagen's (1983) also has a genealogy, as portions of it are developments of his (1974) and (1975).

[5] This difficulty is emphasized by Peter Strawson (1962, 45), who counts himself among "those who don't understand what determinism is."

[6] For what are perhaps the best contemporary treatments of determinism, see Hoefer (2010) and Earman (1986).

and the laws of nature together determine our behavior is the force of *entailment*. What this means, according to the consequence argument, is that if determinism is true, then for me to do otherwise than I actually do, I have to somehow falsify either the past history of the world or the laws of nature. And since falsifying something presumably requires having some sort of power over that thing, it follows that power over whether I raise my cup requires power over the past or the laws—a power which seems impossible. Thus, if determinism is true, neither I nor anybody else has power over whether I raise my cup (or perform any other action). If determinism is true, none of anybody's actions are ever free.[5] The consequence argument, as we saw above, is thus an argument for incompatibilism about causal determinism and free will.

Now you might think that the force of this threat is weakened because we don't know that determinism is true. But even though we don't know that determinism is true, unfortunately we also don't know that determinism is *false*. Even worse, it might not even be *possible* to establish that determinism is false. Concerns about our ability to establish whether determinism is true or false come from several directions (cf. Hoefer 2010, especially §3), but one problem has to do with the laws of nature. It seems that we need to have a pretty good grasp—a better grasp than we currently do—on what the laws of nature *are* before we're able to decide the question of whether determinism is true. In fact, some authors (e.g., Cartwright 1999) have denied the *existence* of laws of nature, muddying the waters even further. In light of these and other considerations, the truth of determinism remains an epistemic possibility. For all we know, determinism could be true, and thus for all we know, none of our actions are ever free.

This, then, is the challenge that I propose to address. Of course, this challenge has been addressed at length in the literature. But some responses have been explored in more detail than others, and I wish to explore a type of response that has, until recently, been largely neglected.

1.3 A parallel

The type of response that I'm interested in is one that construes the consequence argument as running in parallel to an influential version of the *skeptical argument* in epistemology. This type of response to the consequence argument draws inspiration from epistemological responses to skepticism. There are of course different varieties of skepticism, but the variety I'm concerned with is built on the argument that we don't know what we think we know about the external world. (We will examine both arguments in much more detail below.)

While this approach is not entirely new, I don't think that the parallel between these two arguments has been adequately explored. But there have been some helpful pointers in the right direction. For example, Watson (2003) joins others in describing the incompatibilist about human freedom and determinism who affirms determinism (and thus denies freedom) as a skeptic. And Fischer points out that these two arguments—the skeptical argument and the consequence argument—"are similar to the extent that they both challenge deep and widely-held views by employing ingredients that are intuitive and natural":

> The situation here is similar to the challenge to our intuitive belief that we know various things about the empirical world. It is natural to believe that we can have this sort of knowledge. And yet we can naturally be led to question this belief. That is, ingredients that capture widely held and intuitive ideas can be employed to call into question our common-sense view that we can have knowledge of the empirical world. ... These kinds of skeptical arguments issue in a sort of internal tension or "cognitive dissonance," and they challenge us to scrutinize our beliefs more finely. (Fischer 1995, 11)

In both cases we have an intuitively plausible belief that can be challenged in an intuitive way—i.e., challenged by premises that are intuitively plausible in their own right. (Another way to put the point, following DeRose, is to say that these are arguments in which plausible premises yield a conclusion that is highly *im*plausible (1995, 2); the skeptic purports to establish his conclusion by, as it were, using our own beliefs and intuitions against us (cf. 1995, 49).)

But I think the similarities between the skeptical argument and the consequence argument run even deeper than this. For example, both arguments begin with a skeptical hypothesis. The argument that we don't have knowledge of the external world begins with (or at least might begin with) with the hypothesis that all of our perceptual inputs are illusory because we're actually plugged into "the Matrix," which is feeding us sensations (and nutrients) through a tube.[7] This hypothesis is a contemporary version of similar hypotheses that have been used to call into question our knowledge of the external world. Descartes, for example, proposed two such hypotheses: that we might now be dreaming, even though we seem to be awake; and that we might now be under the deceptive influence of an extremely powerful evil genius. A more recent hypothesis suggests that we might not be the embodied persons that we think we are, but rather disembodied brains, floating in nutrient-filled vats while electrodes provide illusory experiences via electrical stimulation. The crucial similarity between these different hypotheses is that we don't believe them to obtain; and yet, were they to obtain, it's possible that our experiences would be phenomenologically indistinguishable from our current experiences (cf. DeRose 1995, 672). And the same is true of the Matrix hypothesis.

According to the Matrix hypothesis, I might *think* that I'm a graduate student studying philosophy, who's been married for eight years and has two children, and who, say, had pancakes for breakfast this morning. But in reality I've lived my entire life in a pod—studying nothing, interacting with no one, and certainly not eating pancakes for breakfast this or any other morning. And, again, the unfortunate thing about this hypothesis (and skeptical hypotheses in general) is that I'm unable to rule it out. For it seems that if I were plugged into the Matrix, the evidence provided by my sensory experiences could be exactly the same as the sensory evidence that I'm getting in a non-Matrix world. The non-Matrix world and the Matrix world appear to be indistinguishable from the inside.

[7] *The Matrix*, of course, is the 1999 movie in which intelligent machines have taken over the world and subjected humans to this treatment so that they can be used as an energy source.

Similarly, the consequence argument begins with the hypothesis of universal causal determinism: the hypothesis that every event that occurs—including our allegedly free actions—is part of the unfolding of the initial conditions of the universe according to the laws of nature. And, as we saw above, the hypothesis of determinism is also a hypothesis that we can't rule out.

By themselves, of course, these hypotheses aren't especially disturbing. It is perhaps *mildly* disturbing that we can't rule these possibilities out, but merely considering them isn't enough to cause us to doubt our knowledge of the external world, or our sense that we sometimes act freely (and moreover that we are often morally responsible for what we do). What is needed, in order to produce the disturbing results, is some premise or set of premises that serves as a bridge from the skeptical hypothesis to the skeptical conclusion (cf. P. J. Graham 2007, 20). Thus the real trouble comes when we consider the *implications* of these hypotheses, given the validity of certain *closure principles*. Therefore, if we want to understand the similarities between the skeptical argument and the consequence argument, we need to examine these closure principles in more detail.

1.4 Closure principles

We will call the closure principle that leads to skepticism about the external world an *epistemic* closure principle. The basic idea behind the epistemic closure principle is that we can extend our knowledge by accepting what's entailed by what we already know. In other words, knowledge is closed under entailment: If I know that *p,* and if *p* entails *q,* then I know that *q*. This principle is invalid in its unqualified form, however, because for some *p,* I might not *recognize* that *p* entails *q*. Thus we need to qualify the principle. There are other qualifications that are arguably necessary as well, and various nuanced ways of making those qualifications, but for present purposes the following revised principle should suffice:

(1) If *S* knows that *p,* and knows that *p* entails *q,* then *S* knows that *q*.

Going forward, we will use (1) to represent the idea that knowledge is closed under (known) entailment. The validity of (1) (and related, perhaps more nuanced epistemic closure principles) is admittedly a matter of controversy—but rather than attempt a defense of any particular formulation, I will simply point out that (1) is intuitive, and that granting it only strengthens the skeptic's position. Thus I will assume that (1) is valid and move on to ask how it might be used to cause trouble for our putative knowledge of the external world.

Recall the skeptical hypothesis considered above, namely that I'm plugged into the Matrix, which is providing me with illusory sensory experiences. Let *p* be the proposition that I had pancakes for breakfast this morning, and let *q* be the proposition that I'm *not* plugged in to the Matrix. With *p* and *q* thus defined, the argument runs as follows:

(2) If I know that I had pancakes for breakfast, and that my having pancakes for breakfast entails my not being plugged into the Matrix, then I know that I'm not plugged into the Matrix. (1)

(3) I know that my having pancakes for breakfast this morning entails that I'm not plugged into the Matrix. (Premise)

So, (4) If I know that I had pancakes for breakfast this morning, then I know that I'm not plugged in to the Matrix. (2, 3)

(5) But I don't know that I'm not plugged in to the Matrix. (Premise)

So, (6) I don't know that I had pancakes for breakfast this morning. (4, 5)

One virtue of this formulation is that it makes explicit the role that the epistemic closure principle plays in the skeptical argument. (And of course there's nothing special about this choice of *p*; the same argument could be run for almost any proposition, no matter how mundane, that I think I know.) Thus we can see that the skeptical hypothesis that I might be plugged into the Matrix, together with the principle that knowledge is closed under known entailment, generates the skeptical worry that I don't know anywhere near as much about the external world as I thought I did.

Notice now that although a principle such as (1) is commonly understood as a closure principle, it could also be understood as a *transfer* principle. This is because knowledge is in a sense being transferred across the entailment from *p* to *q*. Notice also that we can turn (1), which is an epistemic closure or transfer principle, into a different kind of closure or transfer principle by substituting a different modal operator. And we can assess the plausibility of this new principle by asking how plausible it is to claim that the modality in question is transferred across the entailment. One transfer principle that has been discussed in the literature on free will and moral responsibility (e.g., Fischer 1995) is the *transfer of powerlessness* principle. The basic idea behind this principle, roughly speaking, is that if we have no power over something, and no power over whether that first thing leads to a second thing, then we have no power over the second thing. Powerlessness is transferred across entailment. We can encapsulate this transfer principle as follows:

(7) If *S* has no power over whether *p,* and no power over whether *p* entails *q,* then *S* has no power over whether *q*.

The status of (7) seems to be roughly equivalent to the status of (1). It's certainly not uncontroversial, but it is intuitive—and granting its truth only strengthens the case for incompatibilism. As a brief defense of the principle, however, notice that it mirrors the distribution axiom of modal logic. The distribution axiom tells us that if it is necessary that if *p* then *q,* then if necessarily *p* then necessarily *q*. Similarly, the transfer of powerlessness principle is telling us that if it is *power* necessary that if *p* then *q*—if *S* has no power over whether the conditional holds—then if *S* has no power over whether *p* then *S* has no power over whether *q*. And notice that the same can be said for the transfer of knowledge principle, which involves, as Fischer (1995) has called it, *epistemic* necessity. What the epistemic principle is telling us is that if it is epistemically necessary that if *p* then *q,* then if it's epistemically necessary that *p* then it's epistemically necessary that *q*. It's certainly not obvious that power necessity and epistemic necessity have the same modal properties as

necessity *simpliciter,* but it seems at least intuitively plausible to claim that they do. Thus we have additional reason to grant that (1), and now (7), are true, and I propose that we do so.

As we saw above, (1) is what added the bite to the skeptical hypothesis that we're all just plugged into the Matrix. And it seems to me that (7) is what adds the bite to the hypothesis that determinism is true. Letting p be the proposition that represents the complete state of the world (which comprises the entire past history of the world and the laws of nature) at t_0, and letting q be the proposition that I raise my coffee cup to take a drink at t_1, the argument runs as follows:

(8) If I have no power over the past and the laws (i.e., the complete state of the world at t_0), and no power over the fact that the past and the laws together entail that I raise my coffee cup at t_0, then I have no power over whether I raise my coffee cup at t_1. (7)

(9) I have no power over the fact that the past and the laws entail that I raise my coffee cup (at t_1). (Premise)

(10) Therefore, if I have no power over the past or the laws, then I have no power over whether I raise my coffee cup. (8, 9)

(11) But I have no power over the past or the laws. (Premise)

(12) Therefore, I have no power over whether I raise my coffee cup. (10, 11)

I have three preliminary comments on this formulation of the argument. First, like the skeptical argument, it generalizes smoothly: any allegedly free action (performed by anyone), if substituted for q, can be challenged in the same way this argument challenges my raising of my coffee cup. Second, like the formulation of the skeptical argument above, this formulation emphasizes the role played by the transfer principle. Finally, notice that this argument is a challenge to our alternative possibilities. If I have no power over whether I raise my cup, then—given that I raise my cup—I can't refrain from raising my cup. My raising my cup is not up to me, and therefore not free.

1.5 Conclusion

In this introductory chapter we have seen that there are certain striking and under-appreciated parallels between the skeptical argument and the consequence argument. They both use intuitively plausible claims to challenge certain common-sense beliefs that we have about ourselves and our capacities. And they share a similar structure: they combine a "skeptical" hypothesis with a closure principle to produce the disturbing result that we don't have knowledge of the external world, or that we never act freely. This parallel suggests an exploration of whether responses to one of the arguments can be modified to apply to the other argument. But before we undertake that exploration, we need to take a look at some of the more traditional responses to the respective arguments. In the next two chapters, we will look at two powerful but arguably unsuccessful responses to the consequence argument.

• Chapter 2 •

Backtracking compatibilism

2.1 Introduction

Having introduced the consequence argument in the previous chapter, I would now like to examine it in more detail. Here it is again in premise form:

(1) If I have no power over the past and the laws (i.e., the complete state of the world at t_0), and no power over whether the past and the laws together entail that I raise my coffee cup at t_0, then I have no power over whether I raise my coffee cup at t_1. (Epistemic closure principle)

(2) I have no power over whether the past and the laws entail that I raise my coffee cup (at t_1). (Premise)

So, (3) If I have no power over the past or the laws, then I have no power over whether I raise my coffee cup. (1, 2)

(4) But I have no power over the past or the laws. (Premise)

So, (5) I have no power over whether I raise my coffee cup. (3, 4)

As we saw in the first chapter, the first premise—the transfer of powerlessness principle—is disputable, but probably correct. And since granting its truth doesn't make my argumentative task any easier, I have proposed that we grant it. The second premise, moreover, seems indisputable. If it were false, then it would be up to me whether the past and the laws entail that I perform some action. In other words, I could through some action of mine determine whether or not the past and the laws entail some action of mine. But it seems clear that I cannot do this; it seems clear that there's nothing I can do such that if I

were to do it, then in virtue of my doing that thing the past and the laws would entail (or fail to entail) some action of mine.

What we're left with, then, is premise (4). The question of whether, and in what sense, we have power over the past and the laws has been much disputed (as we will see below), and that's where we'll be focusing our efforts in this chapter and the next. As John Perry (2004) points out (along with many others, but Perry puts the point nicely), there are basically two ways for the compatibilist to respond to the consequence argument: by adopting a weaker conception of the laws of nature, or by adopting a weaker conception of ability. I will discuss the first strategy in the remainder of this chapter, and the second strategy in the next chapter. But first, I would like to consider a preemptive response to the consequence argument—namely that it begs the question.

2.2 Does the consequence argument beg the question?

As we have seen, the consequence argument is a way to get from the (epistemically possible) hypothesis that determinism is true (and hence that all future events are entailed by the past and the laws) to the troubling conclusion that there is no free will—that none of us can ever do otherwise than we actually do. In order to understand this preliminary challenge to the argument, namely that it begs the question against the compatibilist, let's take a step back and view the argument from a different angle. If determinism threatens our freedom in the way suggested by the consequence argument, then it is worth asking what would have to be different in order for us to act freely. The problem with determinism, so to speak, is that it doesn't allow us any "wiggle room" (or "elbow room," to use Dennett's (1984) evocative phrase). The difference between a deterministic world and an indeterministic world (i.e., one in which we can do otherwise) is that in an indeterministic world, there is at least one point at which we can perform one of two (or more) options—*without having to change the past or the laws*. Or, as Ginet puts it (in his 1990, and as developed in Fischer 1995), having free will is having the power to add to the given past, consistent with the laws of nature; it's the power

to extend the actual past in more than one way. And it is this intuitive idea, which is implicit in the consequence argument, that has been criticized by some as begging the question.

The relationship between the consequence argument and the idea that our freedom is the freedom to extend the given past, consistent with the laws, is explicit in the exchange between Peter van Inwagen (1975, 2004) and David Lewis (1986). Lewis recasts the consequence argument as a *reductio*. (And since van Inwagen doesn't object to this construal in his response, I won't either.) Lewis's construal of the argument is worth looking at for various reasons, but I'm particularly interested in his construal because it will help lay the groundwork for this chapter and the next. Here's the argument as presented by Lewis:

> I did not raise my hand; suppose for *reductio* that I could have raised my hand, although determinism is true. Then it follows, given four premises that I cannot question, that I could have rendered false the conjunction *HL* of a certain historical proposition *H* about the state of the world before my birth and a certain law proposition *L*. If so, then I could have rendered *L* false. (Premise 5.) But I could not have rendered *L* false. (Premise 6.) This refutes our supposition. (Lewis 1986, 296)

Lewis responds to this argument by pointing out there are two ways in which someone might be able to render a proposition false (i.e., falsify a proposition). "An event would falsify a proposition," according to Lewis, "iff, necessarily, if that event occurs then that proposition is false" (Lewis 1986, 297). I am able to falsify a proposition in the *weak* sense if I am able to do something such that the proposition would have been falsified. I am able to falsify a proposition in the *strong* sense if I am able to do something such that my action, or some event caused by my action, would itself falsify the proposition. So the crucial phrase in the argument above, "could have rendered false," can be given a weak interpretation or a strong interpretation. The problem (for the argument) is that on a consistent reading of this phrase, the compatibilist can argue that Premises 5 and 6 can't both be true. If we take the weak reading, says the compatibilist, then Premise 5 is true but Premise 6 is false. If on the other hand we take the strong reading, then Premise 6 is true but Premise 5 is false.

Lewis's insight, then, was that we need to distinguish between two claims that the compatibilist might make: the strong claim that he is able to break a law of nature, and the

weaker claim that he is able to do something such that, if he did it, a law of nature would be broken. Once we make this distinction, then it becomes clear that the consequence argument fails as a *reductio*, because the crucial premises cannot both be true. (Or at the very least this is a tenable claim for the compatibilist to make.)

Van Inwagen's (2004) response to Lewis's move is, among other things, to advocate a definition of "could have rendered false" (actually "is able to render false") that differs from either of Lewis's definitions (i.e., the weak definition and the strong definition).[8] Here is van Inwagen's definition:

> An agent was able to render a proposition false if and only if he was able to arrange things in a certain way, such that his doing so, *together with the whole truth about the past,* strictly imply the falsity of the proposition. (van Inwagen 2004, 346, emphasis mine)[9]

Here we have a clear statement of the supposition that our freedom is the freedom to add to the given past. It is also now clear how this supposition is related to the consequence argument: the notion of ability on which the argument relies is one according to which we must hold the actual past fixed when evaluating ability claims.[10] The challenge we are considering, then, is the accusation that the consequence argument, insofar as it relies on this notion of ability, begs the question against the compatibilist.

This is a serious accusation, leveled against what seems to be a natural and intuitive idea. (When we deliberate, typically we consider the past as given, and think of ourselves as deciding which of the alternative futures we want to append to the past.) One way to

[8] Actually, as van Inwagen points out (2004, 345n19), this definition was formulated as a result of conversations with Mark Heller, and thus "was not constructed to block Lewis's argument." But it remains true that the revised definition is part of van Inwagen's response.

[9] Horgan (1985) considers this definition at length, explaining how it is an improvement on both van Inwagen's (1977) original definition and Lewis's (1986) gloss. Horgan also offers a further refinement of the definition, and then argues that even the optimal definition of *can render false* will not serve van Inwagen's incompatibilist purposes.

[10] This notion forms the basis for the "basic version" of the argument for incompatibilism, as presented in Fischer (1995, Ch. 5).

interpret this charge is as the claim that the argument begs the question because one of its premises (i.e., "Premise 6" above, when interpreted according to van Inwagen's definition of being able to render a proposition false) is inconsistent with compatibilism. In other words, no committed compatibilist can accept it, because it is part of a valid argument that leads to incompatibilism. But surely being part of a valid argument that leads to incompatibilism does not suffice for begging the question against the compatibilist. This is philosophical disagreement, not begging the question.

Another, slightly more subtle interpretation of this accusation runs as follows. Perhaps this notion begs the question because nobody *would accept the premise* unless they were already convinced of the conclusion (i.e., convinced that incompatibilism was true). Or, to put this interpretation in Maier's (2008, 81) terms, we can determine whether this notion begs the question by evaluating an analysis of freedom that incorporates it. Thus Maier proposes that we evaluate the following schema:

> (6) *S* is free to *A* just in case *S* is free to arrange things in a way such that the whole truth about the past is no different *and S* does *A*.[11]

For Maier, the question is whether (6) is *transparently true*—where (6) can be transparently true only if it is an true *a priori,* and moreover clearly true even to those who are agnostic on the compatibility question. The idea is that if (6) is not transparently true, then the notion that it utilizes (i.e., the notion that our freedom is the freedom to extend the given past) is in danger of begging the question. Although this interpretation of the accusation is an improvement, it suffers from two problems. First, it is not at all clear whether (6) is transparently true, and there is no clear way of deciding the question. So this interpretation doesn't seem to give us a clear verdict. Second, cashing things out in terms of whether

[11] Maier labels the schema "(FS)" rather than "(6)."

someone (e.g., an agnostic on the issue) would accept a premise, or whether some claim would be clearly true to someone, seems to psychologize things a bit too much.[12]

In light of these problems, I suggest that we make a further tweak to our interpretation of the accusation . We should move from the question of whether someone would accept the premise (or the schema) to the question of whether *there are any good reasons* for accepting the premise that are independent of the conclusion. On this interpretation the charge is that any argument that utilizes the relevant notion of freedom begs the question because there are no reasons for accepting it that are independent of the incompatibilist conclusion. And it's true that it's difficult to think of an independent argument for the idea that our freedom is the freedom to extend the given past, consistent with the laws. Such an argument would not only have to be capable of convincing someone who was agnostic about whether freedom was compatible with determinism, but would also have to be capable of doing so without at the same time convincing the agnostic to become an incompatibilist. It's possible that such an argument exists, but I'm certainly not aware of one.

One final way to shed light on the accusation is inspired by Fischer's (1995, 83–85) notion of a "dialectical stalemate." Dialectical stalemates are situations in which one party to the debate (call him *S*) is arguing for some claim *c* on the basis of some principle *p,* which he supports by adducing certain considerations. *S*'s opponent, however (call him *T*), rejects *S*'s argument for *c*—because, he claims, the examples adduced do not support *p* but instead support only a weaker principle *p**. And, continues *T, p** cannot establish *c.* This situation represents a dialectical stalemate because it seems that neither party can move forward without begging the question against his opponent. Examples that decisively establish *p* will most likely beg the question against *T*, and examples that decisively refute *p* will most likely beg the question against *S*. And this, unfortunately, seems to be the position we now find ourselves in. We are considering two positions—compatibilism and incompatibilism—and one principle: that our freedom is the freedom to extend the given past, consistent with the

[12] Thanks to Neal Tognazzini for this point.

laws. And notice that the dialectical situation surrounding the consequence argument appears to be describable in the following way. Anyone arguing for compatibilism is going to have to deny the principle (and claim that any examples or arguments adduced do not support the principle, but some weaker principle instead), but it's not clear how the compatibilist could argue against the principle directly without begging the question against the incompatibilist. On the other hand, anyone arguing for incompatibilism (again, on the basis of the consequence argument) is going to have to affirm the principle—but it's not clear what additional argumentation can be offered in support of it without begging the question against the compatibilist. This may or may not be a situation in which the incompatibilist is begging the question, but it does appear to be a dialectical stalemate in which *both* sides are coming perilously close to doing so. In any case, it is clearly less than ideal.

Given the precarious nature of this situation, it does seem advisable for the incompatibilist to look for an argument that doesn't rely on the premise that our freedom is the freedom to add to the given past, consistent with the laws. I don't have any such argument to offer on behalf of the incompatibilist, but I have tried to construe the consequence argument in a way that avoids this principle. And even if the argument I'm considering does rely on such a principle, I'm going to charitably assume that no begging of the question has occurred, and attempt to respond to the argument on that assumption.

Having considered this preliminary challenge to the consequence argument, we are now in a position to consider two responses that take the argument on its own terms. Development and critique of the first response will take up the remainder of this chapter; the next chapter will be devoted to the second response.

2.3 Backtracking compatibilism

Recall the action that we considered in the previous chapter: I raise my coffee cup to take a drink. Compatibilism, as we have seen, is the view that my freely raising the cup is consistent with that action being determined by the actual past, together with the laws of nature. And on the assumption that acting freely requires the ability to do otherwise, my freely raising

the cup requires my having the power to refrain from raising it. Given this assumption, there are different ways of developing the compatibilist's consistency claim.

Backtracking compatibilists say that we can, without being inconsistent, conjoin certain can-claims, such as

(7) I can refrain from raising the cup,

with certain backtracking counterfactuals—such as

(8) If I had refrained from raising the cup, then some past state of affairs that actually obtained (which, together with the laws of nature entailed that I would raise the cup) would not have obtained.

Backtracking compatibilism, then, is the view that backtracking counterfactuals are sometimes consistent with can-claims.[13]

Alternatively, *local miracle compatibilists* say that we can, without being inconsistent, conjoin (7) with certain local miracle counterfactuals, such as

(9) If I had refrained from raising the cup, then a law of nature (at the actual world) would not have been a law of nature.

The proponent of local miracle compatibilism argues that sometimes agents are able to do something such that an actual law of nature would not have been a law—that local miracle counterfactuals, or "counterlegals," are sometimes consistent with can-claims. (Incompatibilists, of course, will claim that the truth of either (8) or (9) entails the falsity of (7).) I will address local miracle compatibilism in the next chapter; here I would like to develop and defend—although without ultimately endorsing—backtracking compatibilism.

[13] This view is also sometimes referred to as "multiple-pasts compatibilism."

My treatment will largely follow Perry's "Compatibilist Options" (2004), which is in my view the most plausible version of backtracking compatibilism.

According to Perry, the compatibilist who is faced with the consequence argument has three options: adopt a weak account of the laws of nature, or a weak account of ability, or both. (I will focus, as Perry does, on the first two options.) In order to help us understand the difference between these accounts, Perry defines some terms. First he distinguishes (2004, 235) between a proposition's *being* true and its being *made* true. A proposition about a certain time can be true without yet being made true by events. Thus the property of being true isn't relative to times, whereas the property of being made true is. Next, Perry (2004, 235) introduces the notion of *establishing whether p*: "Events *establish whether p* if they make *p* true or make it false." And of course some truths (e.g., mathematical truths) are not made true by any events; Perry points out that for such truths, events *reflect* their truth rather than establish it. And finally, when (and only when) a proposition is entailed by other propositions that have already been made true (some of which have been made true by events, and some of which might have been made true by something other than events), Perry labels it *settled*. Thus the truth of a proposition can be settled before that proposition is made true. The relevant example here, of course, is a proposition *p* about a future action of mine. If determinism is true, then *p* is entailed by a proposition about the past (which was made true by events) together with a proposition about the laws (which was made true by something other than events).[14] Propositions like *p* are settled but not yet made true. Or at least this is what we might call the common-sense view (insofar as common sense applies to the question of how to categorize propositions about future actions, given the truth of determinism).

Perry points out, however, that the commonsensical way of categorizing such propositions as settled is not the only way to categorize these propositions. There is a weaker theory of laws, according to which propositions about my future actions are *not*

[14] Here I am taking the "proposition about the laws" to be a conjunction whose conjuncts together represent all the laws of nature.

settled. Here is how Perry distinguishes between the weak and strong accounts of the laws of nature:

> Is the truth of laws *established* by the events that confirm them and fail to disconfirm them, so that laws are laws because events conform to them? Or is the truth of laws established by something else, so that events *conform* to them because they are laws? The first view is a *weak* theory of laws, the second a *strong* theory of laws. (Perry 2004, 237)

The laws of nature, on the weak theory, are merely true (albeit exceptionless) generalizations, which are partially established by future events (including my future actions). On this view, even if we can say that the past and the laws somehow determine a future action of mine (e.g., my raising of my coffee cup), that action is not yet settled—because, recall, a proposition about a future is settled only when it is entailed by propositions that have already been made true. And in this case, the propositions about the laws have not yet been made true.

The incompatibilist, recall, challenges our freedom by claiming (most likely on the basis of the consequence argument) that if a counterlegal such as

(9) If I had refrained from raising the cup, then a law of nature (at the actual world) would not have been a law of nature

is true, then the corresponding can-claim—in this case,

(7) I can refrain from raising the cup—

must be false. If the compatibilist endorses the weak theory of laws, then it's easy to argue, *contra* the incompatibilist, that (7) and (9) are consistent. This response to the incompatibilist's challenge appears to refute the consequence argument—but this victory comes at a cost. It appears to refute the argument because (9) simply follows from the assumption of determinism and the weak theory of laws —and since events are what make

the laws true, there is no reason to deny the truth of (7). My raising the cup partially establishes the relevant laws of nature at the actual world, but if I had refrained, then that action would have partially established a different law of nature at a different world. These, again, are just implications of the weak theory of laws. This apparent victory comes at a cost because to endorse the weak theory of laws is, after all, to deny the common-sense view: it goes against common sense to say that events establish, rather than conform to, the laws of nature. It is intuitive to view the laws of nature as providing a *constraint* on our behavior, and this theory does violence to that intuitive view. For example, it is a law of nature that nothing can travel faster than the speed of light; and it seems too shallow to say that this is a law merely because nobody has yet caused anything to travel faster than the speed of light. As Perry puts it (2004, 240), "It seems to me much more plausible that this law *gets* at something (or some things) about the universe that explains why things conform to the law and it has no disconfirming instances."[15]

I agree with Perry that it seems more plausible to say that the laws get at something deep about the universe: that events conform to, rather than establish, the laws. But I should note that there have been several recent attempts (e.g., Loewer 1996 and Beebee 2000) to defend a weak, Humean conception of the laws of nature. In fact, Beebee and Mele (2002) have mounted a spirited defense of "Humean compatibilism," according to which the consequence argument fails because there is after all a sense in which I have power over the laws. Hence, according to the Humean about laws, premise (4)

(4) I have no power over the past or the laws

in the above formulation of the consequence argument is false.[16]

[15] See Perry (2004, 240–241) for a brief articulation of the strong theory of laws.

[16] One interesting feature of Beebee and Mele's (2002) treatment of Humean compatibilism is that what they identify as a serious problem for the view—the problem of luck—is also one of the most serious problems for the libertarian.

If a Humean theory of the laws were tenable, then that would certainly be a boon to the compatibilist—for it would provide a straightforward and apparently decisive response to the consequence argument. And while I don't want to rule out the ultimate viability of such a project, I would like to affirm the intuitive and apparent metaphysical force of the laws if possible. Thus I will join Perry, at least for now, in rejecting a weak theory of the laws and searching elsewhere for a response to the incompatibilist challenge.

2.4 A weak account of ability

As mentioned above, the other compatibilist option is to maintain a strong account of the laws but endorse a weak account of *ability*. Here is how Perry introduces the distinction between a weak account of ability and a strong account of ability:

> Can one have the ability to perform or refrain from an action *A* at time *t,* even though the issue of whether one will perform *A* at *t* or refrain from doing so has been *settled* before *t*? A weak account of ability will allow us to answer *yes* to this question; a strong account will force us to answer *no*. (Perry 2004, 237)

So the question of which theory of ability one is dealing with is the question of whether one's theory is committed to the following principle:

(10) If *S* can perform *A* at *t,* then at no time earlier than *t* is it settled whether *S* performs *A* at *t*.

A strong theory of ability will include a commitment to (10), whereas a weak theory of ability will not. Now recall that a proposition is settled if and only if it is entailed by other propositions that have been made true (either by events, or by something else). Recall also the backtracker that we considered above:

(8) If I had refrained from raising the cup, then some past state of affairs that actually obtained (which, together with the laws of nature entailed that I would raise the cup) would not have obtained.

The relevant question at this point is whether (8) is consistent with the can-claim (7). On the strong theory of the laws (which is what we're operating with, having rejected the weak theory in the previous section), they are made true by something other than events. Given the truth of determinism, the proposition that I raise my cup is entailed by the past and the laws, and therefore settled. And if (10) is true, then I cannot refrain from raising my cup at *t*—because at *t* it has already been settled whether I raise my cup.

If, however, we're operating with a weak account of ability, then we will reject (10): *S* can perform *A* at *t,* even though it is already settled whether she will. But why should we adopt a weak account of ability? Is it as implausible as a weak theory of laws? The answer is no: in this case common sense is on the side of the weak. Perry (2004, 241–42) provides a nice analogy to illustrate this point. It involves Elwood, who, in 1956, has an extreme aversion to Edsels and as a result doesn't buy one. This context includes the following feature, which Perry calls a law of nature:

(11) Reasonable people don't buy cars that they think are ugly, ungainly, and overpriced and that they simply don't want and have no other reasons to buy.[17]

The context also includes this feature, which is a fact about Elwood's mind:

(12) Elwood thinks Edsels are ugly, ungainly, and overpriced, and has no other reason to buy one.

From these premises, we are licensed to infer the following fact about Elwood's action(s):

[17] This argument comes directly from Perry (2004), but I have renumbered the premises.

(13) Elwood won't buy an Edsel.

Perry points out, however, that we *aren't* licensed to make the following inference about Elwood's ability to purchase the Edsel:

(14) Elwood can't afford an Edsel.

This last step is invalid because it's a conclusion about Elwood's *finances,* whereas the premises only involve facts about Elwood's *mind*; and there is no connection between facts about Elwood's mind and facts about Elwood's finances. Let *t* be the time at which Elwood is considering the question of whether to buy an Edsel. It is settled at *t* that Elwood won't buy an Edsel, because his not buying an Edsel is entailed by (11), which is a law of nature, and (12), which has been made true by past events. But the question of whether Elwood can *afford* an Edsel is an independent question, and it very well may be that Elwood *can* afford an Edsel, despite its being settled that he won't buy one. (For example, Elwood might be wealthy.)

With this analogy in mind, we have a template for applying the weak account of ability to the question of whether I'm able to refrain from raising my coffee cup to take a drink. If we can pry apart the question of whether it's settled that I'll take a drink from the question of whether I'm able to refrain from taking a drink, then there is a case to be made that I *am* able to refrain from taking the drink. In the case of Elwood and the Edsel, it's easy to separate the question of whether it's settled that he won't buy the Edsel from the question of whether he can afford to buy it: answers to those two questions can obviously come apart. (Although of course there are various pecuniary ways in which it could be settled that he won't make the purchase; for example, he could go bankrupt prior to his deliberations.)

Can we similarly separate the relevant questions as they pertain to the metaphysics of agency? This, I think, is the crucial question, and I will be returning to it in the next

chapter. As a preliminary point, however, notice that the question of whether Elwood can afford an Edsel is a question that will be determined by (what we might call) our theory of financial ability. This is perhaps an extravagant way to put it, but the point remains: if we're wondering whether someone can afford something, we typically ask whether he has enough money in his checking account, or enough room on his credit cards, or some related question. Each of these questions represents an appeal to a condition that is taken to be sufficient for being able to afford the relevant item. I don't have a position on which theory of financial ability is the right theory, but whatever it is, its conditions have nothing to do with Elwood's mental states—which, in Perry's example, are what settle the issue of whether he's going to buy the Edsel. I also don't have a (firm) position on which theory of ability in general is the correct theory, but if its conditions don't explicitly involve the past or the laws of nature, then it's plausible to think that the question of whether, e.g., I can refrain from raising my coffee cup is independent of the question of whether the past and the laws determine that I raise my cup. Consider Perry's proposed account of ability:

> A person has the ability to *bring it about that R in circumstance K* if (i) the person's repertoire of basic actions includes some movement *M* such that (ii) executing *M* in *K* will have the result that *R*. (Perry 2004, 245)

These conditions, continues Perry, "clearly can be satisfied even if the person's not executing *M* falls under a law of nature to the effect that a person with his motivating complexes will not execute *M*." Even though it is settled that the person *won't* execute *M*, it may still be true that she *can* execute *M*.

To summarize: Perry's position is that when I raise my coffee cup to take a drink, it may (and likely will) be the case that I can also refrain from raising my cup—even if determinism is true. In this scenario (in which determinism is true and I raise my cup but could have refrained), the following backtracking counterfactual will be true:

(8) If I had refrained from raising the cup, then some past state of affairs that actually obtained (which, together with the laws of nature entailed that I would raise the cup) would not have obtained.

However, the following counterlegal counterfactual will be false:

(9) If I had refrained from raising the cup, then a law of nature (at the actual world) would not have been a law of nature.

Perry holds, in short, that the can-claim (7) is consistent with a backtracker, but not a counterlegal; hence he is a backtracking compatibilist. I will now consider some objections to this variety of backtracking compatibilism.

2.5 Why not endorse backtracking compatibilism?

Recall that Perry identifies two primary options for the compatibilist: adopt a weak theory of the laws of nature, or adopt a weak account of ability. A weak theory of the laws conflicts with common sense, which leads Perry to endorse and develop the second option. But even though this second option allows a more commonsensical theory of the laws, there's at least one way in which a weak account of ability also conflicts with common sense. As we saw above, Perry's endorsement of a strong theory of laws commits him to the inconsistency of a can-claim and its corresponding counterlegal counterfactual. Instead, he endorses a backtracking counterfactual: if Elwood were to do otherwise, then the past would have been different. This makes him a backtracking compatibilist. But now let us ask just how far back into the past we have to go to reach the point at which the alternative past differs from the actual past. It turns out that if causal determinism is true, then we have to go *all the way back*. Perry makes this point, and expresses some discomfort about it. Supposing that he has chosen to refrain from taking a cup of water on an airplane, this is what he has to say about the alternative past in which he does take the water:

> If I had taken the drink, freely and voluntarily, then surely my beliefs and preferences would have been different than they actually were. The most likely difference would be that I was thirsty. Assuming determinism, if I had been thirsty when the drink was offered, then something earlier also would have been different; perhaps I wouldn't have taken a drink at the fountain before stepping on the plane, as I did, because the fountain was broken. And that would mean some earlier state of the fountain and its surroundings had been different. And so on. Tracing the changes back to the big bang, perhaps it might be a slight difference in the direction in which one particle began its travels through time. Or perhaps it goes back to a deistic god creating the initial state of the universe a very little bit differently. Or perhaps it just goes back, infinitely. Who knows? It's certainly amazing and weird and in my opinion somewhat depressing that the trail of differences that would have led to my being thirsty rather than not being thirsty should lead back even a couple of thousand years, much less to the beginning of time, or forever. Still, I can't see why [something contrary to the laws of nature would have to happen] for me to take the glass. (Perry 2004, 251–52)

I agree with Perry here: it's amazing and weird, and perhaps even depressing, that my doing otherwise would require changes that go back so far into the past. It doesn't *seem* like this is what's required in order for me to do otherwise, even if determinism is true. Common sense, it seems, would suggest something else. My claim, then, is that backtracking compatibilism is just as counterintuitive as local miracle compatibilism. The only difference is that they are counterintuitive at different points. And from a broader perspective this shouldn't be surprising. Local miracle compatibilism is characterized by its commitment to the possible consistency of a counterlegal and its corresponding can-claim. One of the implications of this commitment is a counterintuitive theory of the laws. Backtracking compatibilism, on the other hand, is characterized by its commitment to the possible consistency of a backtracker and its corresponding can-claim. One of the implications of this commitment is a counterintuitive position on what would have to be different, were I to do otherwise in a deterministic world. Backtracking compatibilism may be the correct view, but I don't think the issue can be decided on the basis of which view has common sense on its side.

In the previous paragraph I conceded that one of the implications of the local miracle compatibilist's commitment is a counterintuitive theory of the laws of nature. But I'm not so sure that I'm willing to make this concession, because it seems to me that the key move made by the local miracle compatibilist—the claim that if I were to do otherwise, then

an actual law of nature would not have been a law—is consistent with various positions on the laws of nature. The local miracle view is certainly amenable to a regularity theory of the laws, according to which the laws don't carry any metaphysical weight, but the view doesn't *require* such a theory of the laws. If determinism is true, then, let us grant, my actions are constrained by the laws of nature in some way. And if the local miracle compatibilist is right, then were I to do otherwise, different laws would have been in effect—but my alternative action would have been constrained by those alternative laws in precisely the same sense that my actions in the actual world are constrained by the actual laws.

In the next chapter, I'll attempt to develop the local miracle view in more detail (and in a way that is neutral with respect to different accounts of the laws of nature). For now, though, let's look at two more problems with backtracking compatibilism, which have been raised by Fischer (1995, 2008).

The first problem has to do with practical reasoning, and can be motivated by the example of the icy patch:

> Sam saw a boy slip and fall on an icy patch on Sam's sidewalk on Monday. The boy was seriously injured, and this disturbed Sam deeply. On Tuesday, Sam must decide whether to go ice-skating. Suppose that Sam's character is such that if he were to decide to go ice-skating at noon on Tuesday, then the boy would not have slipped and hurt himself on Monday. (Fischer 1995, 95)

According to the backtracking compatibilist, the relevant can-claim—

(15) Sam can decide to go ice skating at noon on Tuesday—

is true, as is the relevant backtracker:

(16) If Sam were to decide to go ice-skating on Tuesday, the accident wouldn't have occurred on Monday.

But now consider Sam's deliberations. Given the truth of the relevant backtracker (16), it seems that Sam *ought* to decide to go ice skating—because if he were to decide that then the accident wouldn't have occurred. On the other hand, Sam knows that the accident *did* occur, so a decision to go skating also seems *irrational*; it seems like wishful thinking. So we have a bit of a puzzle. Allowing for the truth of the relevant backtracker, consistent with the truth of the can-claim, preserves Sam's freedom but at the cost of introducing irrational wishful thinking into his practical reasoning.

Notice also that there's nothing particularly special about this example. If backtracking compatibilism is true, then there would seem to be pressure toward this kind of wishful thinking any time an agent is able to perform some action, and if she were to perform it then some unfortunate state of affairs in the past would have turned out better than it in fact did.[18] Thus the backtracking compatibilist seems to allow various considerations into our practical reasoning that, intuitively, should not be allowed.

One way to solve this puzzle, of course, is to insist that the backtracker is false. Perhaps there's a different conditional, involving a lapse of memory or character in the consequent, rather than a difference in the past. But the plausibility of these different conditionals is going to depend on the details of the story, and there are certainly some stories in which it will be clear that the backtracker is the true conditional. Therefore, it's better to solve the problem in a way that allows for the truth of the backtracker.

Here is the solution that Fischer offers. First, restrict the reasons that are relevant for deliberation to those reasons "that obtain in the possible worlds one can actualize (or which are accessible to one)"—where the only possible worlds that one can actualize are those worlds which share a past with the actual world (Fischer 1995, 95). In other words, the only accessible worlds are those in which the future is an extension of the actual past. If Sam follows this policy in his deliberations, then he won't be led into thinking that he ought to go

[18] In fact, some other examples from Fischer (1995, 80–81)—the solicitous friend, the careful historian, and the salty old seadog—indicate that wishful thinking is not the only kind of irrationality that would afflict our practical reasoning if backtracking compatibilism is true.

skating in order to prevent the accident—because the only reasons he'll consider are reasons that obtain in light of the accident's already occurring.

This does solve the puzzle in a way that's neutral with respect to the truth of the backtracker, but notice that it requires the controversial notion of extending the actual past. And even if this notion doesn't beg the question, it is fatal to compatibilism. So it seems that the backtracking compatibilist is left with a dilemma. If he accepts this solution, then he has solved the problem at the expense of his overall position—he's won the battle but lost the war. If, on the other hand, he rejects the solution, he's stuck with the question of how to reject the problematic pattern of practical reasoning that threatens to lead Sam (for example) into irrational decisions.

Even if this dilemma is escapable, the backtracking compatibilist would still face what Fischer (2008) calls the problem of "baggage." The basic idea behind this problem can be summarized as follows. First, recall the crucial question that Perry raises: Can we do otherwise than we actually do, even if it's already been settled that we won't? (And recall that a proposition is settled if it's entailed by propositions that have already been made true.) His answer is that we can, as long as the relevant proposition, which is settled, hasn't yet been made true by events. If this is right, then I can, e.g., refrain from raising my cup—even though the raising is entailed by past events and the laws of nature—so long as the proposition that I raise my cup hasn't yet been made true. And if I were to refrain from raising my cup, then there is a true backtracker: some proposition about the past, which "be" true in the actual world, would not have been true. According to Perry, if we distinguish between (a) an ability to act so that some true proposition would not have been true and (b) an ability to cause a true proposition to be false, then it's clear that we can have the ability described in (a) even if the ability described in (b) is impossible.

Fischer (2008) argues that this distinction doesn't make a difference. To understand his point, consider an explicit statement of two of the relevant propositions, along with the corresponding backtracking counterfactual:

(17) I raised my cup at t_3.

(18) I felt sleepy at t_1.

(19) If I had refrained from raising my cup at t_3, then I wouldn't have felt sleepy at t_1.

Let us suppose that, in the actual world, (17)–(19) are true. (Strictly speaking, of course, (19) might not be the true backtracker; it might not be (18) that would have been false, had I refrained, but some other proposition(s) about the past instead. To simplify exposition, I'll suppose that (18) is the relevant proposition about the recent past and thus that (19) is true.) Perry's claim, again, is that even though (17) might have been settled prior to my act of cup-raising (settled at, say, t_2)—because it was entailed by propositions about the past (including (18)) and the laws of nature—it hadn't yet been made true. Moreover, I could have refrained from raising my cup, and had I refrained, then (18) would have been false. In other words, in the nearest possible world in which I refrain from raising the cup, both (17) and (18) are false.

Fischer points out (or would point out, if he were discussing the example of my raising the cup) that, at t_2, even though (17) hasn't yet been made true by events, (18) has. So it follows from my ability to refrain from raising the cup that I am able to act such that an already-made-true proposition, namely (18), would have been false. This is what we learn from the truth of (19), and this is why propositions such as (17) come with baggage: they are entailed by propositions which have already been made true—propositions which would have to have been false, had the agent done otherwise than he actually did. The problem of baggage is, according to Fischer, the reason why the distinction between a proposition's being made true by events and its being merely settled fails to rescue the compatibilist from the consequence argument.

To sum up: We have seen that backtracking compatibilism allows for an intuitive view of the laws of nature, but a counterintuitive view of what would have to be different, were I to do otherwise in a deterministic world. In this respect it appears to be on a par with

local miracle compatibilism. We have also seen that there are some puzzles of practical reasoning, not to mention a baggage problem, that afflict the backtracking compatibilist. Even though it's not entirely clear how serious this affliction is, these considerations should motivate us to at least explore the other compatibilist option: local miracle compatibilism.

• Chapter 3 •

Local miracle compatibilism

3.1 Introduction

As we saw in the previous chapter, backtracking compatibilism is the view that backtracking counterfactuals are sometimes consistent with can-claims. Backtracking compatibilists, that is, claim that we can, without being inconsistent, conjoin certain can-claims, such as

(1) I can refrain from raising the cup,

with certain backtracking counterfactuals—such as

(2) If I had refrained from raising the cup, then some past state of affairs that actually obtained (which, together with the laws of nature entailed that I would raise the cup) would not have obtained.

Alternatively, *local miracle compatibilists* say that we can, without being inconsistent, conjoin (1) with certain local miracle counterfactuals, such as

(3) If I had refrained from raising the cup, then a law of nature (at the actual world) would not have been a law of nature.

The proponent of local miracle compatibilism argues that sometimes agents are able to do something such that an actual law of nature would not have been a law—that local miracle counterfactuals, or "counterlegals," are sometimes consistent with can-claims. (And

incompatibilists, of course, will claim that the truth of either (2) or (3) entails the falsity of (1).)

The previous chapter's discussion motivated an exploration of local miracle compatibilism. I would like to carry out this exploration in light of an influential challenge to the local miracle view, as presented by Carl Ginet in his seminal book, *On Action* (Ginet 1990). Ginet's challenge, roughly speaking, is that when local miracle compatibilists give up the entailment between the truth of counterlegals and the falsity of can-claims, their account loses much of its explanatory power. For local miracle compatibilists must acknowledge that there are some cases in which the relevant counterlegal is true, and yet the corresponding can-claim is false (for example, cases in which the can-claim involves an action that the agent obviously cannot do). But the compatibilist apparently cannot explain the agent's inability in these cases, because he cannot appeal to the truth of the counterlegal as the reason why the can-claim is false. The upshot of my response to this challenge is that in these cases (in which a counterlegal is true, but the corresponding can-claim is not) the compatibilist should use his preferred analysis of 'can' to explain the falsity of the can-claim.

My aim in this chapter is to develop a version of local miracle compatibilism along these lines. I will stop short of fully endorsing local miracle compatibilism, in part because the view itself will remain incomplete: I will not, for example, be providing the analysis of ability that would be required, were I to develop the view in its entirety. My defense of the view will however emphasize and further support the idea, introduced in the previous chapter, that we can (and should) separate questions about the truth of ability claims from questions about whether causal determinism obtains.

After summarizing the main claims of local miracle compatibilism, I will consider a common objection to the view. My response to this objection will set the stage for Ginet's challenge, which I will consider at length. I will then offer and defend a response to Ginet's

argument on behalf of the local miracle compatibilist.[19] I will close by summarizing the state of the dialectic thus far.

3.2 Local miracle compatibilism (and its detractors)

I begin with some general remarks about local miracle compatibilism. This view is notable because it provides a distinctive answer to the question of which counterfactuals would be true, if causal determinism were to obtain (and yet we could do otherwise than we actually do). So, contrary to backtracking compatibilists, local miracle compatibilists claim that typically (though not invariably), the counterlegal conditionals would be true.[20]

Another way of describing local miracle compatibilism, as we have seen, is to say that its proponents are committed to the claim that counterfactuals such as (3) are consistent with can-claims such as (1). This commitment, however, is not equivalent to the claim that *only* counterlegals (and thus not backtracking counterfactuals) are consistent with can-claims.[21] Nor is it equivalent to the claim that counterlegals are *always* consistent with can-claims. Instead, the local miracle compatibilist maintains (or at least should maintain) that in any given case, it is an open question whether (and which) counterfactuals are true. Moreover, he maintains that local miracle compatibilism does not in itself have any general answer to this question.

Failure to understand this aspect of the local miracle view leads some of its critics into thinking that defenders of the view are attempting to provide both necessary and sufficient conditions for being able to do otherwise. But typically they are not. This becomes clear if we return to Lewis (1981). Lewis's defense of local miracle compatibilism begins with

[19] Another recent defense of local miracle compatibilism (written in response to a different and more recent complaint) can be found in P. A. Graham (2008).

[20] In addition to Perry (2004, 2008), Jonathan Bennett (1984) is another example of a backtracking compatibilist.

[21] It is possible, of course, to hold a more restrictive, or "pure," version of local miracle compatibilism according to which only counterlegals are consistent with can-claims (and hence the truth of backtrackers would rule out freedom). Such purism seems to me unnecessary and implausible. Below I say more about the benefits of a "hybrid" view.

intuitions about what we can and cannot do. We can raise our arms (when everything is in working order) but we cannot jump over buildings. He then moves from these intuitions, taking them for granted, and says that if determinism is true, then it follows that certain counterlegal conditionals are true. But not to worry—those conditionals do not require attributing to the agent the incredible ability to cause lawbreaking events. In the context of this defense, Lewis is clearly not taking himself to be providing an informative account of what exactly we are able to do or of which counterfactuals are true and which are false. Instead, he is merely defending his compatibilism against an attack from the consequence argument. Thus, to interpret his view (and local miracle compatibilism in general) as providing an exhaustive account is to misconstrue the dialectical situation.[22]

With these preliminary points in mind, let us examine what is perhaps the most common criticism of local miracle compatibilism, namely that its proponents are committed to saying fantastic things. For example, van Inwagen claims that it is incoherent to challenge the limits that laws of nature impose on our abilities:

> Suppose a bureaucrat of the future orders an engineer to build a spaceship capable of traveling faster than light. The engineer tells the bureaucrat that it's a law of nature that nothing travels faster than light. The bureaucrat concedes this difficulty, but counsels perseverance: "I'm sure," he says, "that if you work hard and are very clever, you'll find some way to go faster than light, even though it's a law of nature that nothing does." Clearly his demand is simply incoherent. (van Inwagen 1983, 62)

If the local miracle compatibilist's commitment to the consistency of (3) and (1) is relevantly similar to the bureaucrat's counsel above, then the compatibilist's view would be fantastic indeed. But of course the local miracle compatibilist will deny the similarity between his commitment and the bureaucrat's demand. On what basis is the local miracle compatibilist entitled to this denial? Well, the laws of nature entail that nothing can travel faster than light. If the engineer were to carry out the bureaucrat's demand, then he would have to falsify this fact that nothing can travel faster than light—i.e., he would have to build

[22] Thanks to Michael Nelson for this way of formulating the point.

something that travels faster than light. And nobody can falsify something that is entailed by the laws of nature. (Following Ginet (1990, 112), I will say that an event falsifies *p* if and only if its occurrence is sufficient for the falsity of *p*.) Thus, the reason the bureaucrat's demand is incoherent is that, for it to be carried out, the engineer would have to do the impossible: falsify something that is entailed by the laws of nature. The bureaucrat's demand is inappropriate because the content of the demand is something that no one can do.

These considerations can be brought together in the following principle:

(4) If *p* is entailed by the laws of nature, then it is never open to anyone to perform an action that would be or cause an event that falsifies *p*.[23]

What (4) tells us, then, is that no agent can perform an action that would be or cause a lawbreaking event. This principle is important for at least two reasons. First, it explains why the bureaucrat in van Inwagen's example is in no position to demand that the engineer build a spaceship that can travel faster than light: given our laws, satisfying such a demand is inconsistent with (4). Second, this principle illustrates a crucial disanalogy between the bureaucrat's demand and the local miracle compatibilist's claim that (3) and (1) are consistent. The key difference between the two is that the bureaucrat's demand—but not the compatibilist's claim—requires that the agent in question do something that would be or cause a lawbreaking event. Building a spaceship that travels faster than light would cause a lawbreaking event, whereas refraining from raising a coffee cup would not be (and would not cause) a lawbreaking event. This is because, according to local miracle compatibilism, had I refrained from raising my cup, the "miracle" would have occurred just prior to the cup-raising.[24] Hence, the bureaucrat's demand, but not the local miracle compatibilist's claim,

[23] See Ginet (1990), although the general strategy comes from Fischer (1995), who is following in the footsteps of Lewis.

[24] This, at least, is the standard view about when the local miracle occurs. Vihvelin (2000) takes a different approach according to which the miracle occurs at the same time as the relevant choice.

runs afoul of (4). The proponents of local miracle compatibilism can defend themselves against the charge that they are committed to fantastic results by adopting (4).

3.3 Ginet's argument against the local miracle view

Despite its apparent usefulness as a response to van Inwagen's bureaucrat case, Ginet is not satisfied with (4); he thinks it is too weak to capture "our conception of the limits placed on our freedom by the laws of nature (1990, 113)". And of course any compatibilist view, such as the local miracle view, that is built on an inadequate conception of the ways in which the laws of nature constrain our freedom will itself be inadequate. Ginet thinks (4) is too weak because it is not able to license certain commonsense inferences from premises involving antecedent conditions and laws of nature to conclusions about what we as agents can and cannot do (the latter class of actions being especially important). The principle encapsulated in (4) licenses such inferences in a wide range of cases (e.g., in van Inwagen's bureaucrat case), but Ginet (1990, 113–14) argues that there are certain cases in which it cannot do the work that it needs to do. In support of this claim, he provides the following case:

> Suppose that some time before *t, S* ingested a drug that quickly causes a period of complete unconsciousness that lasts for several hours. Suppose that, because of the drug, there is true of *S* a certain proposition of the form
>
> At *t, S*'s neural system was in state *U*
>
> and it follows from this proposition and the laws of nature that *S* was unconscious for at least thirty seconds after *t*.

Now consider the following propositions:

(5) At *t*, *S*'s neural system was in state *U*.

(6) Beginning at *t* plus five seconds, *S* voluntarily exerted force with her right arm for ten seconds.[25]

[25] These propositions are taken, though numbered differently, from Ginet (1990, 114).

According to Ginet (1990, 114), given the details of the case, we should be able to conclude that (6) is false: "we are surely entitled to deduce that it was not open to *S* to voluntarily exert force with her arm in the five seconds after *t*." But, as Ginet points out, (4) fails to license this inference. To see why, first assume that the limitations that the laws of nature place on our actions are in fact encapsulated by (4)—which, recall, can be rendered as

> (4) If *p* is entailed by the laws of nature, then it is never open to anyone to perform an action that would be or cause an event that falsifies *p*.

On this assumption, in order for us to be able to infer that it is not open to *S* to voluntarily exert (i.e., raise) her arm at *t*+5, we have to suppose that she does raise her arm at *t*+5, and then show how that event, in itself, falsifies some proposition *p entailed by the laws of nature*. The relevant *p* in Ginet's example is this: "If *S*'s neural system is in state *U* at *t,* then it is not the case that she raises her arm at *t*+5." But *S*'s voluntary exertion at *t*+5 does not falsify this *p,* because her raising her arm then does not entail that she was in state *U* at *t*. In other words, we can imagine a possible scenario in which *S* raises her arm at *t*+5, but was not in state *U* at *t*. Thus, her raising her arm does not entail the falsity of the conditional *p* because it is consistent with the falsity of *p*'s antecedent. Moreover, we cannot infer (at least not from (4) alone) that it was not open to *S* to raise her arm at *t*+5.

Perhaps Ginet's argument will become clearer if we look at some of the other inferences that might be made, based on the details of his case. Consider two action descriptions, A_1 and A_2. Let A_1 be *S raises her arm at t+5 while in state U,* and let A_2 be *S raises her arm at t+5*. The incompatibilist claims that we can only perform actions that extend the actual past, consistent with the actual laws. Given the details of the case, the incompatibilist can infer that *S* is unable to perform the actions described by both A_1 and A_2. (This is because, as the story goes, *S* is as a matter of fact in state *U* at *t*+5, so there is no extension of the actual past in which she raises her arm at that time.) The compatibilist, however, does

not require that free actions be an extension of the actual past, consistent with the laws. (And recall that even the local miracle compatibilist will allow for the truth of the occasional backtracker—a point I will have more to say about below.) The compatibilist *can* infer that *S* is unable to perform the action described by A_1, since her performing that action would falsify *p* ("If *S*'s neural system is in state *U* at *t,* then it is not the case that she raises her arm at *t*+5"), which is entailed by the laws of nature. But the compatibilist *cannot* infer that *S* is unable to perform the action described by A_2, because there is no *p* (that is entailed by the laws of nature) that such an action would falsify. The incompatibilist, then, has a rule of inference that allows him to conclude, based on the details of the case, that *S* is unable to perform the action described by A_2; the compatibilist's adoption of (4) invalidates this rule of inference. Less formally, we might say that the compatibilist has opened up some "wiggle room" by endorsing (4) rather than the stricter incompatibilist criterion.

Ginet's challenge can be summarized as follows: From the details of the case he presents, along with certain principles about the past and the laws, we should be able to infer that *S* was not able to voluntarily move her arm at *t*+5—i.e., that *S* does not have it within her power to so act that (6) would be true. But the principle in (4) cannot deliver this result, because the laws of nature do not entail, by themselves, that *S* does not raise her arm for 10 seconds. Thus, there is nothing involved in her raising her arm that would be or cause an event that falsifies something that is entailed by the laws of nature, and as a result we have no explanation of *S*'s inability to raise her arm. In what follows, I will defend the local miracle compatibilist against this powerful challenge.

3.4 In defense of local miracle compatibilism

The gist of my response to Ginet's challenge is that the local miracle compatibilist's principle about the relationship between the laws of nature and our ability (or inability) to do otherwise need not provide, in and of itself, an explanation of our inability to do otherwise in *every* scenario in which we cannot do otherwise. This is because the compatibilist has another principle or set of principles, derived from his analysis of 'can,' or his criteria for the

truth of can-claims, which rules out doing otherwise in certain circumstances—and in some cases this latter principle is what does the explanatory work. (What I am doing, in other words, is roughly gesturing toward the informative account that Lewis was not interested in providing.)

Recall that local miracle compatibilists are notable for arguing that, given mere causal determination, counterlegal conditionals such as the following are often true:

(3) If I had refrained from raising the cup, then a law of nature (at the actual world) would not have been a law of nature.

Moreover, in some cases the truth of a conditional such as (3) does not rule out a corresponding can-claim, such as

(1) I can refrain from raising the cup.

Finally, recall that the details of various situations in which both determinism and a can-claim like (1) are true—and not some prior commitment of the local miracle compatibilist—will determine which counterfactuals are true in those situations. With these points in mind, the first step in my defense of local miracle compatibilists is simply to note that their basic claim is consistent with an acknowledgement that there are certain cases in which a backtracking counterfactual could be true as well. The local miracle compatibilist can insist that (3) and (1) could both be true, while also allowing that it might be true that

(2) If I had refrained from raising the cup, then some past state of affairs that actually obtained (which, together with the laws of nature entailed that I would raise the cup) would not have obtained.

In other words, the local miracle compatibilist—unlike the backtracking compatibilist—can allow that there are some situations in which both a counterlegal and a backtracker will be

true (although in such a circumstance presumably the backtracker will not trace back indefinitely into the past).[26]

The second step in my defense is the claim that there may be cases in which both a counterlegal and a backtracker will be true, but the corresponding can-claim will *not* be true. However, in such cases the proponent of local miracle compatibilism should not say that the truth of either conditional, in itself, rules out the can-claim. Rather, the local miracle compatibilist should invoke his preferred analysis of 'can' (or his criteria for when can-claims are true) and point out that the obtaining of some related condition or circumstance (e.g., the agent's being unconscious) is what renders the can-claim false.

The second step leads us to the third and most important element of my proposed defense. Ginet's challenge, recall, is that the local miracle compatibilist's principle (4), which details the relationship between the laws of nature and our (in)ability to do otherwise, does not explain why we are licensed to infer that *S* is not able to raise her arm at *t*+5. I propose that the appropriate response to this challenge is to point out that, for the local miracle compatibilist, the status of a principle such as (4) is similar to the status of the relevant counterlegals and backtrackers. In certain situations, the relevant can-claim is false and there are true counterlegals and backtrackers; but in these situations, it is not necessarily the truth of these conditionals that rules out the truth of the can-claim. It may be that in these situations the compatibilist's analysis of 'can' is what explains the falsity of the relevant can-claim. Similarly, the local miracle compatibilist can say that in certain situations, the relevant can-claim is false and this is explained by (4)—because for the can-claim to be true, the act in question would have to falsify something that is entailed by a law of nature. In other situations, the relevant can-claim is false and this is explained by his analysis of 'can,' or his criteria for the truth of 'can' claims. In other words, the compatibilist's principle (4) need not

[26] There is a stronger claim that we could make here, namely that certain local miracle strategies, such as Lewis's, *require* the truth of at least some backtrackers, because the lawbreaking event that makes some agent's action possible will have to have occurred before the agent's willing—which of course means that had the agent done otherwise, the past would have been different. For more on counterlegals and backtrackers, see Lewis (1979b).

explain, in and of itself, the falsity of can-claims in every scenario—because he has another principle (derived from his analysis of 'can'), which details the circumstances in which we are unable to do otherwise than we actually do, and in certain cases this latter principle is what does the explanatory work.

My proposal becomes clearer if we take a closer look at the role that (4) plays in the dialectic surrounding local miracle compatibilism. Recall that the local miracle compatibilist who distinguishes between an ability to break a law and an ability to do something such that, were he to do it, a law would have been broken, is not necessarily attempting to provide an exhaustive account (and hence an explanation) of what we can and cannot do. In other words, (4)—as we saw above—is an essential element of a *defense* against the charge that local miracle compatibilists are committed to fantastic claims. Given its status as part of a defense, it is not surprising that (4), by itself, will often fail to license inferences about what we are unable to do. (In fact, it would be strange to expect it to invariably license such inferences.) The fundamental explanation of why we are unable to do the things that we are unable to do is not going to be in terms of (4), but rather in terms of the compatibilist's analysis of ability.

What I am suggesting, then, is that Ginet's case, in which *S* is unable to raise her arm, is a case in which the compatibilist's analysis of 'can,' rather than a principle such as (4), is what explains the relevant inability. To see how this suggestion might play out, consider first a simple (and no doubt false) conditional analysis of 'can':

(CA) *S* can do *A* just in case if *S* were to choose to do *A,* then *S* would do *A*.

According to this analysis, *S*'s doing *A* is "choice-dependent."[27] (Similarly, proponents of this analysis would presumably want to say that *S*'s being able to do otherwise than *A* consists in her doing otherwise, were she to choose to do otherwise.) But if we apply this analysis to

[27] I have adopted this treatment of the conditional analysis (including its problems and possible revisions) from Fischer (2007, 49–53).

Ginet's case, it should be obvious that it gives us the wrong result. *S* is unable to raise her arm at *t*+5, and we want to be able to explain why she was unable to raise her arm (i.e., could not have done otherwise than she did). But her inability to raise her arm is not the result of a faulty connection between her (hypothetical) choice and her exertion. For it is true that had she chosen to raise her arm, then she would have raised her arm; it is just that she cannot choose to raise her arm at *t*+5 (because she is unconscious). So (CA) tells us that *S* could have done otherwise (because her choosing to raise her arm is choice-dependent), which is the wrong result.

However, it is open to the proponent of (CA) to acknowledge that, by itself, the truth of the right side of the biconditional in (CA) is not sufficient for the truth of the left side; it is instead merely a necessary condition for *S*'s ability to do *A*. He can then modify (CA) by adding a further necessary condition to the effect that there is no obstruction or barrier preventing *S*'s choosing to do *A*. These two conditions (the original condition and the new condition) would then be individually necessary and jointly sufficient for the truth of "*S* can do *A*." So the revised analysis would look something like this:

(CA′) *S* can do *A* just in case
(1) if *S* were to choose to do *A,* then she would do *A,* and
(2) there is no barrier to *S*'s choosing to do *A*.

If we apply this revised analysis to Ginet's case, then it seems to deliver the right result: it is true (we are assuming) that if *S* were to choose to raise her arm, that she would raise her arm, but it is *not* true that there is no barrier to her choice. There is in fact a fairly obvious barrier to her choice, namely the fact that she is unconscious. So, according to (CA′), *S* is not able to raise her arm at *t*+5.

Unfortunately, even (CA′) suffers from at least three weaknesses. The first weakness is that the notion of "barrier to choice" is too vague. The second is that some barriers to choice are more controversial than others. (For example, one's being unconscious is an

uncontroversial barrier to one's choice, whereas one's being the subject of subliminal advertising, or hypnosis, might be more controversial.) So (CA′) will be of limited use when the choice in question is prevented by a controversial barrier. The third weakness is that (CA′) arguably begs the question against the incompatibilist, and is therefore of limited utility in the debate over whether an appropriately robust sort of freedom is compatible with causal determinism.[28]

Despite these admittedly serious weaknesses, (CA′) is sufficient for my purposes. First, even though the "barrier to choice" notion is vague, the basic idea is intuitive. For my purposes, all the local miracle compatibilist needs is the possibility of a suitably refined notion and the resultant suitably refined analysis. (And I see no reason to think that this refined notion is *impossible*.) Second, notice that at this stage in the dialectic the local miracle compatibilist is not trying to cover all cases in which there are barriers to choice. Nor, thirdly, is he trying to settle the issue of whether freedom is compatible with causal determinism. (Although he is presupposing *some* defense of the compatibilist thesis.) Rather, he is simply trying to point out that his view is capable of dealing with Ginet's case—which is to say that his view can deliver the result that *S* is not able to raise her arm, given that she is unconscious. His view says that *S* was unable to do otherwise in Ginet's case because there was an obstacle to her choice: the fact that she was unconscious. An analysis of 'can' such as (CA′) will admittedly not help establish the truth of compatibilism. But it need not, so long as there is another aspect of the compatibilist's view that can do that job.

Note also that my defense of the local miracle compatibilist does not force him to adopt a conditional analysis of 'can'; we could press the same point with a different compatibilist analysis. The local miracle compatibilist could, for example, adopt Keith Lehrer's (1976) possible worlds approach, according to which 'can' is not analyzed in terms of

[28] (CA′) arguably begs the question against the incompatibilist because the incompatibilist will no doubt insist that some action's being entailed by the past and the laws is a "barrier" to an agent's being able to refrain from that action. This point comes from Fischer (2007, 51–52).

conditionals, but rather in terms of accessibility relations between the actual world and various possible worlds. Although his analysis is complex, and worthy of a fuller treatment than I will give it here, a rough sketch should suffice to show that the plausibility of local miracle compatibilism does not depend on the plausibility of any conditional analysis of 'can."[29]

Lehrer's first point is that the conditional analysis falls prey to various counterexamples; hence the need to seek out a superior analysis. His alternative proposal begins with a possible worlds semantics according to which "minimal difference" (rather than the comparative similarity approach advocated by Stalnaker and Lewis) is used to evaluate subjunctive conditionals.[30] Lehrer's basic claim is that whether *S* can do *A* depends on "what sort of minimal difference would be required" to accommodate *S*'s doing *A* in some possible world (1976, 240). An important additional element of Lehrer's analysis is that the minimally different possible world in terms of which *S*'s abilities are evaluated (call it w_1) cannot be one in which *S* has an unfair (or "inadmissible") advantage, relative to the actual world w_0 (in which *S* does not do *A*). Lehrer offers a precise definition of which sorts of advantages are admissible,[31] but the intuitive idea is that an advantage is admissible only if it can be gained through *normal* means between the time t_1 at which the can-claim is evaluated and the time t_2

[29] Fischer (1979) offers a fuller treatment of Lehrer's article, as well as an extended criticism.

[30] For Lehrer, minimal difference is evaluated while holding fixed the laws of nature. But this is not an essential part of his view, and so could be dropped by the local miracle compatibilist, were he to adopt Lehrer's analysis of ability. For examples of approaches to evaluating minimal difference between worlds that do not hold the laws fixed, see Audi (1978), Horgan (1977), and Horgan (1979).

[31] Lehrer's analysis, including his definition of an admissible advantage, can be found in Lehrer (1976, 256–257): " '*S* could (at t_i) have done *A* at t_n' is true in the actual world *W* if and only if there is a possible world *w* having the same laws as *W* and minimally different from *W* so that '*S* does *A* at t_n' is true in *w* in such a way that any advantage *S* has in *w* for doing *A* at t_n which he lacks in *W* is admissible for *S* from *W* and t_n is past. An advantage *S* lacks in *W* is admissible for *S* from *W* if and only if either (a) the advantage results from *S* doing something *B* at t_j ($t_i \leq t_j \leq t_n$) when he has no additional advantage for doing *B* at t_j in *w* which he lacks in *W* or (b) the advantage results from *S* doing something *C* at t_k ($t_i < t_k \leq t_n$) when *S* has no additional advantages for doing *C* at t_k in *w* which he lacks in *W* except those advantages admissible to *S* from *W* resulting from what *S* does prior to t_k."

at which the action is performed. Of course, "normal" is not exactly a precise term, and would need to be replaced with something more precise, were I considering adopting Lehrer's analysis as part of an overarching compatibilist theory. But I am merely trying to illustrate how Lehrer's analysis might be applied to a case such as Ginet's, and for that purpose the intuitive notion of a normal way of gaining an advantage should suffice.

So, for example, this morning (as I write this) it is true that I can drive to the airport this afternoon. Doing so requires some advantages that I do not have now (e.g., being in my car), but these are the sorts of advantages that I can gain through normal means (e.g., by walking out to my car, opening the door, and getting in). So even though I do not actually go to the airport this afternoon, the minimal difference required in a possible world in which I do go to the airport is one that does not include any inadmissible advantages. On the other hand, if it is now 10 minutes before I am supposed to be at the airport, and it normally takes 30 minutes to drive there, then the minimal difference between w_0 and w_1 would involve me having an inadmissible advantage (e.g., the freeways being completely empty between here and the airport, and a car that will go three times the speed limit). This advantage is inadmissible because there is no way for me to bring it about (using normal means); something completely and radically fortuitous would have to happen for me to make it to the airport in 10 minutes. Hence, such an advantage is not an admissible advantage, and Lehrer's analysis tells us that it is false (10 minutes before I am supposed to be at the airport) that I can make it to the airport on time.[32]

In summary: Lehrer analyzes can-claims by saying that *S* can do *A* in w_0 if and only if *S* does *A* in some possible world w_1 that is accessible to *S*. For w_1 to be accessible to *S,* it must be "minimally different" than w_0, and that minimal difference must not include any inadmissible advantages. It seems clear that this analysis will also deliver the result that *S,* in Ginet's example, is unable to raise her arm at *t*+5. This is because there do not seem to be any possible worlds that are accessible to *S* in which she raises her arm at *t*+5. The only

[32] The airport example comes from Ginet (1990, 111). See note 18 for a discussion of his version.

worlds in which she raises her arm are going to be worlds in which she has an inadmissible advantage, and hence worlds that are not minimally different (and so not accessible) from the actual world.

There are, admittedly, legitimate doubts about the ultimate success of Lehrer's account as a reductive analysis of ability.[33] But I am not trying to provide a reductive analysis of ability. A complete defense of local miracle compatibilism would perhaps require commitment to some specific analysis; but I am not trying to provide a complete defense either. Instead, I am seeking merely to defend the view against one important challenge by making a point about the *structure* of local miracle compatibilism. I am pointing out that the structure of local miracle compatibilism is such that compatibilists can appeal to certain additional principles to explain why we are unable to do certain things in certain situations (e.g., the situation Ginet describes, in which *S* is unable to raise her arm). It is true that these principles are hard to nail down, and it is probably also true that no satisfactory account of them has yet been offered. But of course it does not follow that no such account is possible, and I see no reason why a compatibilist should not be confident that some such account exists.[34] For my purpose (i.e., a response to Ginet), all the local miracle compatibilist needs is

[33] See for example Fischer (1979).

[34] Thanks to Peter A. Graham for this way of formulating the point. Graham also points out (in personal correspondence) that the incompatibilist is not really in a better position than the compatibilist, at least when it comes to explaining why certain people are unable to do certain things in certain situations. To understand this point, recall that Ginet's criticism is that the compatibilist cannot appeal to (4) to license an inference to the conclusion that *S* is unable to raise her arm. The incompatibilist is supposed to be in better shape with respect to this inference, since he can appeal to a principle about the inescapability of the laws (i.e., a principle which states that nobody can do anything such that were she to do it, a law would have been broken) to license the inference in question. But Graham provides a slightly modified case in which the inescapability of the laws cannot provide the inability verdict that seems intuitive. The modified case is simply an indeterministic variation on Ginet's. Suppose that *S* takes the drug, which causes her to be unconscious for 30 seconds following *t*. Suppose also that while she is unconscious, a number of indeterministic events happen, one or two of which are such that had they gone the other way (a way that it was physically possible for them to go), *S* would have woken up and raised her arm. It seems intuitive that even in this modified case we would want to say that it is false that *S* is able to raise her arm at *t*+5. Presumably the incompatibilist will agree that *S* is unable to raise her arm in this scenario. But what resources does the incompatibilist have here that will allow him to explain why *S* is unable to raise her arm? He cannot appeal to the inescapability of the laws, because there are a number of physically possible worlds in which

the possibility of an analysis of ability that is able to deliver the appropriate result in Ginet's case (and cases relevantly similar to it).[35]

In short, it seems as though the local miracle compatibilist should respond to Ginet's challenge by saying something resembling the following:

> First, it is obvious that *S* cannot raise her arm (at *t*+5). And if she did raise her arm, then the past would have been different—i.e., she would not have been given the drug that caused her neural system to be in state *U* (or at least it would have failed to cause her neural system to be in state *U*). But it is not the truth of this backtracker, or a principle about the laws of nature, that renders the can-claim false. Instead, the can-claim is false because the conditions of my compatibilist analysis of 'can' are not met—there is a (non-nomological) barrier to *S*'s choice that obtains at the relevant time. (Alternatively, *S*'s not being in *U* at *t*+5 would be an inadmissible advantage that she lacks in the actual world.) Thus, I can infer that *S* is unable to raise her arm (at *t*+5) without making the further claim that the truth of the counterlegal (or the backtracker) rules out the truth of the can-claim; I can say that the truth of the can-claim is ruled out by my analysis of 'can.'

We can now review our progress. First, the local miracle compatibilist says that causal determinism does not, in itself, imply that the conditions for 'can' are not met. Further, we have seen that the compatibilist can adopt (4) as a defensive strategy against the charge that he is committed to fantastic claims about the relationship between can-claims and the laws of nature. We have also seen that the compatibilist has additional resources available to him: in addition to a principle that relates the laws of nature to our inabilities to perform various actions (the principle encapsulated in (4)), he has an analysis of 'can' that delivers plausible inability verdicts in certain scenarios. (This might be a conditional analysis, or a possible worlds analysis, or some other, more satisfying analysis.) In some cases (such as

---

she does raise her hand. What the incompatibilist needs is some principle—apart from the principle of the inescapability of the laws—that rules out these other physically possible worlds (in which *S* raises her arm) as accessible to the agent. Thus, the incompatibilist, contrary to what is implied by Ginet's challenge, is not in a better position to explain why *S,* for example, is unable to raise her arm when intuitively it seems that she cannot.

[35] As I have alluded to above, a local miracle compatibilist might even refrain from adopting any specific analysis of 'can' claims, but rather accept certain constraints on any plausible analysis—where those constraints ensure that the analysis will deliver the appropriate verdict in cases such as Ginet's.

van Inwagen's) an agent (the engineer) may not be able to perform some action, and this might be true because performing that action would require falsifying something that is entailed by the laws of nature. In other cases (such as Ginet's), the agent (*S*) also may not be able to perform some action, but in those cases the agent's inability will be explained by there being an uncontroversial barrier to her choosing to do that action—a condition that obtains that will prevent her from performing the action no matter what analysis of 'can' is adopted. (Note also that because of the presence of this barrier, there will be a backtracker that is true of the scenario: had the agent performed the action, the barrier would have to have been absent.[36])

---

[36] There is evidence that Ginet might be sensitive to the point that I am making here—for example, when he says the following:

> The Local Miracle View can allow for *exceptions* ... . For example, if it is true that Jones was at the faculty meeting at ten minutes before noon and it is true that, if he had been in the airport at noon, then it would have to have been the case that he was not at the faculty meeting at ten minutes before noon, then it must also be true that it was not open to Jones at any time during that ten minutes to make it the case that he was in the airport at noon. (Ginet 1990, 111, my emphasis)

This case is structurally identical to the case in which *S* is unable to raise her arm, as evidenced by the fact that we can straightforwardly plug in the details from that case:

> If it is true that *S* ingested a drug (at *t*) that quickly causes a period of complete unconsciousness that lasts for several hours and it is true that, if she had raised her arm at *t*+5, then it would have to have been the case that she had not ingested the drug at *t*, then it must also be true that it was not open to *S* at any time between *t* and *t*+5 to make it the case that she raised her arm at *t*+5.

(For clarity of exposition, I have changed the time at which *S* ingested the drug from "some time before *t*" to *t*. In other words, I am assuming that the drug acts immediately.)

The modified example is somewhat convoluted, but the point is simple: if *S* took the drug at *t,* then she is not able to raise her arm at *t*+5. Hence, if she does raise her arm at *t*+5, then it must be that she did not take the drug at *t.* Moreover, there is nothing she can do between *t* and *t*+5 that will enable her to raise her arm. The explanation for these facts, as we saw above, is that there is a barrier to *S*'s choosing to raise her arm at *t*+5 (namely, her being unconscious at *t*). If she does raise her arm, then the barrier must not have been in place. Likewise, the explanation for Jones's situation is that there is a barrier to his being at the airport at noon (namely, his being in the faculty meeting ten minutes before noon). If he is at the airport at noon, then the barrier must not have been in place. If an "exception" is allowable in Jones's case, then, given the parallel between his predicament and *S*'s predicament, it is hard to see why it is not allowable in *S*'s case.

The view I am suggesting, then, is a "mixed" or hybrid view insofar as it holds that in any given scenario, counterlegals or backtrackers might be true. But, again, it will not be the truth of these counterfactuals that explains our abilities to do what we are able to do. Similarly, the view I am proposing is one that will, on occasion, appeal to a principle such as (4)—but (4) will not always be what explains our *in*abilities to do what we are *not* able to do. To criticize the local miracle compatibilist for not being able to use (4) to deliver an inability verdict in certain cases—as Ginet does—is to disregard the compatibilist's analysis of ability, and hence to unfairly limit his access to his theoretical resources.

3.5 Ginet's response

Ginet does consider a response like mine, so it is worth examining the way in which he rejects it. He begins as follows:

> The compatibilist may reply that, although the example shows *that we rely on something stronger than (Y)* [my (4)], its apparent demonstration that we rely on the inescapability of the laws is still an illusion. (Ginet 1990, 114, my emphasis)

The principle of the inescapability of the laws is the principle that Ginet favors over (4). This inescapability principle says, roughly, that nobody can do anything such that, were she to do it, a law would have been broken (and thus not a law).[37] The inescapability of the laws is clearly a stronger principle than (4), as it rules out a broader range of actions than (4) does. For example, supposing that I raise my coffee cup in a deterministic world, it rules out my refraining from that raising of my coffee cup, whereas (4) does not. It seems, then, that Ginet interprets the local miracle compatibilist as acknowledging that (4) is too weak, but

[37] More precisely, the principle of the inescapability of the laws says that "if *p* is deducible from the laws of nature, then it is never open to anyone to make it the case that not-*p*" (Ginet 1990, 105). According to Ginet's (1990, 100–101) notion of making it the case that *p*, "*S* made it the case that *p* if and only if *p* and *S* caused (at least) the last thing needed for it to be the case that *p*." Thus, the inescapability principle implicitly holds fixed certain aspects of the actual past—including the actual laws of nature. And if the laws must be held fixed in this way, then I am not able to do anything such that, were I to do it, an actual law would not have been a law.

maintaining that there is a principle out there that is stronger than (4) but still weaker than the inescapability of the laws.

Although it is true, formally speaking, that the compatibilist response I am proposing is in part an acknowledgement that the compatibilist needs something stronger than (4), putting it in those terms obscures the essence of the move I am suggesting. What I am suggesting is that Ginet's example shows that the compatibilist relies on something *in addition* to (4)—namely, his analysis of 'can.' The reason why it is misleading to view this addition as simply a stronger version of (4) is that the compatibilist's analysis of ability, as I point out above, is going to be what provides the fundamental explanation in the vast majority of cases in which some agent *S* is unable to do something *A*. Recall that (4) is merely employed as part of a defensive maneuver, and so should not be viewed as a basic principle that is strengthened in the face of counterexamples. It is, rather, an ancillary principle that is used to fend off the charge that local miracle compatibilists are committed to saying fantastic things.

The remainder of Ginet's response can be summarized as follows (1990, 115–17). What the compatibilist might say (in response to his example) is that *S*'s doing *A* at t_2 was avoidable at t_1, even if *S* was nomically necessitated at t_1 to do *A* at t_2—so long as the necessitation passed through *S*'s prior motivational states and processes in the right sort of way. The compatibilist might further claim that the example in which *S* is unconscious only supports the inescapability of the laws if we import in the additional premise that the necessitation according to which *S* cannot raise her arm does not pass through her motivational states. Without this implicit premise, the example fails to establish the inescapability of the laws.

This compatibilist suggestion does not work, argues Ginet, because freedom to do otherwise is in fact lost if one's actions are nomically necessitated—even if that necessitation involves motives and will in the right sort of way. To see this, imagine what it would be like to know the laws of nature that govern our actions, and to use this knowledge to manipulate someone else (*S*) into performing some action *A*. In this case, we should conclude that *S* is

not able to avoid doing *A*—even if she is being manipulated in a way that utilizes her normal motivational processes. And this conclusion will be supported, in part, by the inescapability of the laws—"as part of a more complex inference of the form given in our [i.e., Ginet's] argument for incompatibilism." (Ginet 1990, 116)[38]

This is a powerful response to what is perhaps the standard compatibilist line, but it fails if considered as a challenge to the defense of compatibilism that I am proposing. This failure can be demonstrated in the form of a dilemma.

Ginet's response here is either intended to rule out what he considers the most plausible compatibilist analysis of ability (i.e., one according to which it is essential to *S*'s freely doing *A* that *A* be the right sort of result of her motives and will), or intended to rule out the possibility of *any* compatibilist analysis of ability. If his response is directed toward a single compatibilist analysis (or family of analyses), then all it has shown is that a compatibilist principle (or set of principles) designed to explain our inability to do otherwise will not deliver the right result in manipulation cases—at least not if the principle is appealing only to motivational states. But this claim from Ginet is consistent with my own claim, namely that the imagined principle (which is, let us grant, unable to deal with manipulation cases) is the *sort* of principle that the compatibilist should appeal to when dealing with Ginet's case in which *S* takes the drug. The introduction of manipulation cases makes it clear that the compatibilist has work to do before he can claim that he has produced necessary and sufficient conditions for our being able to do otherwise. But it does not follow from this that no such compatibilist conditions are possible.

But perhaps—moving now to the second horn of the dilemma—Ginet is intending that his response rule out *all* compatibilist analyses of ability. (His comment about the "more complex inference of the form given in [his] argument for incompatibilism" supports this interpretation.) The problem with this strategy is that it is dialectically inappropriate. The

[38] Ginet then considers and dismisses the soft compatibilist line about there being a difference between necessitation as a result of manipulation and necessitation as a result of natural causes.

only argument (in the current context) that is capable of such a sweeping conclusion is the consequence argument.[39] But, to reiterate, local miracle compatibilism was developed precisely as a response to the consequence argument. Local miracle compatibilism may, in the end, fail—but one cannot simply point to the consequence argument as the reason why it fails. Its embrace of the truth of certain counterlegals moves the dialectic beyond the consequence argument. Moreover, I would add that the possibility of a hybrid version of local miracle compatibilism (according to which certain backtrackers might be true in addition to the true counterlegals, and according to which it is not necessarily the truth of any counterfactual that explains our inability to do otherwise) makes it far from clear that all compatibilist accounts of ability must fail.

3.6 Conclusion

I will close by summarizing what I take to be the state of the dialectic. Those who are critical of local miracle compatibilism tend to think that the local miracle compatibilist is committed to saying that in all cases in which an agent intuitively is not able to do otherwise, that inability can be explained by invoking a principle about the laws of nature (and how they constrain our abilities). But this assumption, as I hope to have shown above, is not true. The distinctive feature of local miracle compatibilism is indeed the willingness of its proponents to affirm local miracle counterfactuals such as (3)

> (3) If I had refrained from raising the cup, then a law of nature (at the actual world) would not have been a law of nature.

Moreover, the truth of these counterfactuals is consistent with some relatively weak principle, such as (4)

[39] Ginet presents his version of the consequence argument in (1990, chapter 5).

(4) If *p* is entailed by the laws of nature, then it is never open to anyone to perform an action that would be or cause an event that falsifies *p,*

which tells us what some proposition's being entailed by a law of nature means for our abilities regarding the truth of that proposition. But both of these features, or components, of the view are employed in the service of a defensive strategy: The local miracle compatibilist says that free will is compatible with determinism; his detractors wonder how that can be, since determinism precludes the ability to do otherwise. The local miracle compatibilist responds with a local miracle counterfactual: we (often) can do otherwise, and if we had done otherwise, then an actual law of nature would not have been a law of nature. His detractors say this is ridiculous, because if it were true then we could, for example, hire someone to build a spaceship that travels faster than the speed of light. The local miracle compatibilist says, on the contrary, that he has a principle relating the laws of nature to our (in)ability to do otherwise, and according to that principle it is not open to anyone to do something as fantastic as build a spaceship that travels faster than light. The detractors then point out that this principle cannot explain our inability to do otherwise in every scenario. The local miracle compatibilist in turn points out that his principle need not provide, in and of itself, such an explanation in every scenario, because he has another principle (derived from his analysis of 'can'), having to do with straightforward (i.e., non-nomological) barriers to our doing otherwise, and in certain cases this latter principle is what does the explanatory work.[40] At each turn, detractors accuse the local miracle compatibilist of saying something outlandish; in response, the compatibilist employs resources for explaining why his commitments are not outlandish after all.

I hope that a careful consideration of Ginet's critique of local miracle compatibilism has helped to make clear that the local miracle compatibilist has more resources available

[40] Moreover, the local miracle compatibilist might continue, this proliferation of principles is not *ad hoc* or otherwise unjustified, because any local miracle account is going to (1) need an analysis of 'can' and (2) want to evaluate the truth of various counterlegals and backtrackers in the most plausible way.

than merely a proposition such as (4), relating power and the laws of nature. Yes, the local miracle compatibilist adopts (4) as a means of defense against the accusations of extravagant or outlandish consequences. But such a compatibilist can avail himself of other resources—such as an analysis of the relevant sort of power or ability, or even just plausible components of such an analysis—in order to explain cases such as the one Ginet offers. Ginet's critique, then, while not fatal to the local miracle view, can be seen as prompting local miracle compatibilists to go at least some way toward developing the informative account that earlier proponents of the view were not interested in providing.

Local miracle compatibilism is thus a nuanced doctrine, involving different layers of ingredients and typically coming in a hybrid form (that allows for some true backtracking counterfactuals along with some true counterlegals). This hybrid nature, I would argue, puts it in a slightly better overall position than backtracking compatibilism—especially in light of the problems with backtracking compatibilism that we considered in the previous chapter. Nevertheless, while it is important to note the complex logical structure of local miracle compatibilism, this is not in itself an argument for the acceptance of the doctrine. As I noted above, a full-blown argument for local miracle compatibilism would require a commitment to some specific analysis of ability. I have argued that the proponent of the local miracle view *can* provide an explanation for our inabilities in a wide range of cases (including cases such as the one Ginet presents); but without endorsing an analysis of ability, I have not actually provided the needed explanation.[41] But I do take myself to have shown—building on the previous chapter, and in preparation for subsequent chapters—that the question of whether the past and the laws of nature have settled that, for example, I raise my coffee cup should be distinguished from the question of whether I can refrain from raising that cup. That is, the answers to these two questions might come apart. The correct answer

[41] For example, in the absence of a compatibilist analysis of ability, one might wonder why being unconscious now *would* be a barrier to doing something a little later, while (in another case) simply not wanting to do the thing *wouldn't* be a barrier—since each state of affairs (in its case) is part of a causally sufficient condition for one's not doing the thing. Thanks to Randy Clarke for emphasizing this point in personal conversation.

to the first question depends on the ultimate success or failure of the consequence argument, whereas the correct answer to the second question depends on our criteria for can-claims.

At this point we have apparently made as much progress as the traditional responses to the consequence argument will allow. In order to make further progress, we need to take a step back and examine the argument that runs in parallel to the consequence argument: the skeptical argument in epistemology.

• Chapter 4 •

The argument for skepticism

4.1 Introduction

Having spent some time on the consequence argument, and on two of the more influential responses to it, this much is now clear: although there may be some hope for the eventual success of either local miracle compatibilism or backtracking compatibilism, there are enough difficulties with these views (at least at their current stage of development) to motivate us to explore other options. This exploration is what will occupy us for the better part of the next three chapters. I will begin by taking a step back and revisiting the parallel (introduced in Chapter 1) between the consequence argument and the skeptical argument in epistemology.

4.2 The skeptical argument

Recall that the skeptical argument,[42] like the consequence argument, contains two main ingredients. The first ingredient is a skeptical hypothesis: an epistemically possible scenario that, for all we know, could be the actual scenario. The consequence argument begins with the deterministic hypothesis. Our world might be deterministic, and for all we know it is deterministic. The skeptical argument could begin with one of any number of hypotheses, but the one I've chosen to work with is what I will call the *Matrix hypothesis*: the possibility that despite the way things seem to us, we are hibernating in pods somewhere while machines stimulate our brains to produce our experiences. If this possibility were actual,

[42] As with the consequence argument, it would be more precise to say that "the skeptical argument" refers to a *family* of arguments—one member of which we will be focusing on in what follows.

then the external world, although it would exist, would be radically different from the way we perceive it to be.

Remember also that these hypotheses, as troubling as they may be, are not by themselves sufficient to challenge our freedom or our knowledge. What's needed in addition to these hypotheses is some sort of closure, or transfer, principle. The consequence argument (or at least the version we're focusing on) relies on a transfer of powerlessness principle, and the skeptical argument relies on a transfer of knowledge principle. Following tradition (and in order to help us distinguish between the parallel principles), I will from now on refer to the transfer of knowledge principle as the principle that knowledge is closed under known entailment—or, for short, simply the epistemic closure principle. Here, again, is the principle:

(1) If *S* knows that *p,* and knows that *p* entails *q,* then *S* knows that *q*.

Below we will consider the viability of a view that denies closure. For now, though, as in Chapter 1, we will maintain our commitment to it on the basis of its intuitive plausibility.

Let *p* be the proposition that I had pancakes for breakfast this morning. Let *q* be the proposition that I'm not plugged into the Matrix. What (1) tells us is that if I know that I had pancakes for breakfast, and if I know that my drinking coffee entails that I'm not plugged into the Matrix, then I know that I'm not plugged into the Matrix. But it's hard to see how I could know that I'm not plugged into the Matrix—since if I were, then my experiences (we can suppose) would be exactly the same as they are now. Therefore I don't after all know that I'm drinking coffee. And this conclusion extends to a myriad of other ordinary propositions that we think we know—and represents a troubling challenge to our common sense beliefs regarding what we know about the external world.

Here is the more formal version of the argument, which was introduced in Chapter 1:

(2) If I know that I had pancakes for breakfast, and that my having pancakes for breakfast entails my not being plugged into the Matrix, then I know that I'm not plugged into the Matrix. (1)

(3) I know that my having pancakes for breakfast this morning entails that I'm not plugged into the Matrix. (Premise)

So, (4) If I know that I had pancakes for breakfast this morning, then I know that I'm not plugged into the Matrix. (2, 3)

(5) But I don't know that I'm not plugged into the Matrix. (Premise)

So, (6) I don't know that I had pancakes for breakfast this morning. (4, 5)

I have set things up so that the structure of the skeptical argument is parallel to the structure of the consequence argument. Recall that there were two primary options for the compatibilist who wants to reject the incompatibilist conclusion of the consequence argument: an alternative account of the laws of nature, or an alternative account of ability. Either of these options, if viable, allows the compatibilist to reject the premise that claims that we have no power over the past and no power over the laws. The anti-skeptical options don't divide up quite as neatly into two (since there's no obvious analogue in the skeptical argument to the role that the past and the laws play in the consequence argument), but for our purposes a two-fold division will suffice. The first option involves relaxing the requirements for knowledge so that certain skeptical possibilities need not be ruled out in order for us to have knowledge. (As we will see, this option might also involve a denial of the epistemic closure principle.) If this relaxed conception of knowledge (often referred to as a "fallibilist" conception) is the correct one, then we can see our way clear to denying (5). The second option that would enable us to deny the skeptical conclusion involves advocating for a *contextualist* position according to which the standards for knowledge attributions vary according to context. On this view (about which more in Chapter 5), which possibilities need to be ruled out in order for a particular belief to count as knowledge will depend on details of the conversational context. In ordinary contexts, it would be false to claim, for example, that I don't know that I had pancakes for breakfast. However, the introduction of a

skeptical scenario (such as the Matrix hypothesis) raises the standards for knowledge attributions by expanding the range of possibilities that need to be ruled out in order for me to know that I had pancakes for breakfast. In other words, according to the contextualist story, the skeptical argument only goes through by relying on the artificially high standards that result from the introduction of the skeptical hypothesis in the first premise.

I will consider the first of these options in this chapter, and the second option in the next chapter. In order to provide a framework for entertaining the relaxed conception of knowledge, it will be helpful to proceed, much as we did in Chapter 2, by asking whether the skeptical argument begs the question against the non-skeptic.

4.3 Does the skeptical argument beg the question?

As we consider the possibility that the skeptical argument begs the question, first recall the (admittedly not entirely satisfactory) notion of begging the question that we came up with in Chapter 2: A premise begs the question when there are no reasons for accepting it that are independent of the conclusion that it supports. Or, to adopt Maier's (2008) helpful terminology, a premise begs the question unless it's "transparently true."

As we saw in Chapter 2, some have accused the consequence argument of begging the question because the notion of ability that it uses is a notion according to which I am only able to do otherwise (than I actually do) if there is a possible world, in which I do otherwise, that shares both the actual past and the actual laws of nature. Given the parallels we've already seen (between the consequence argument and the skeptical argument), it seems worthwhile to ask whether there is a notion that is employed in the skeptical argument that might beg the question against the non-skeptic. And in this case it's not hard to figure out where to start, because there's really only one concept with respect to which the skeptic might be begging the question: the concept of *knowledge*.

If we examine the crucial premise of the skeptical argument—

(4) If I know that I had pancakes for breakfast this morning, then I know that I'm not plugged into the Matrix—

we can see how this accusation of begging the question might get some traction. First, let's take the contrapositive of (4), which gives us the logically equivalent

(7) If I don't know that I'm not plugged into the Matrix, then I don't know that I had pancakes for breakfast this morning.

The idea behind (7), just to reiterate, is that if I can't rule out a certain sort of possibility, then I can't know a certain sort of everyday proposition. This suggests that there is some threshold of measurement—for example, degree of similarity to the actual world—and that scenarios which fall within the relevant measurements need to be ruled out in order for someone to have knowledge of the relevant proposition (which is relevant because it entails the falsity of the relevant scenario). But it is difficult to see how we can draw this line (between scenarios that need to be ruled out and scenarios that don't need to be ruled out) in any principled, non-absolute way. Absent some proposal for such a principle, it seems that we are led by our acceptance of (7) to the claim that we must rule out *all* possibilities, no matter how bizarre or abstruse, in order to have knowledge of some proposition.[43] So the preliminary question to consider as we examine the skeptical argument is this: Are there any reasons to accept this absolutist notion of knowledge that are independent of the skeptical conclusion? And again—as with the consequence argument—it's difficult to think of any such reasons. Why think that I have to rule out all possibilities in order to know that I had pancakes for breakfast? Can't I know what I had for breakfast even though I can't rule out the obtaining of every bizarre and fanciful scenario? But, again, if we are going to limit the possibilities that need to be ruled out, we need to do so in a principled way; this will put us

[43] Perhaps we can say that we "only" need to rule out metaphysical possibilities, and not logical possibilities that are metaphysically impossible, but this hardly helps matters.

on firmer footing when considering the accusation that the skeptical argument begs the question. In the next few sections, we will examine some proposals for the desired principle.

4.4 Fallibilism

The first proposal we will consider—*fallibilism*—is not so much a principle as it is an approach. The proponent of fallibilism rejects the notion that all possibilities of error must be ruled out in order to possess knowledge. That is, he argues that absolute certainty is not required for knowledge. This fallibilist claim is supported by both linguistic and practical considerations.

First, the linguistic considerations: it is not uncommon to claim knowledge—e.g., "I know that the bank will be open on Saturday"—and yet, in the face of a challenge, deny absolute certainty: "Well, no, I'm not *absolutely certain* that the bank will be open on Saturday. It's possible that they changed their hours since the last time I was there." Moreover, at least in our ordinary language, such a denial of absolute certainty does not amount to a retraction of one's knowledge claim (cf. Feldman 2003, 123 and Pritchard 2005, 20). Contrast this with a denial of belief ("I know that the bank will be open on Saturday, but I don't believe that the bank will be open on Saturday"), which seems paradoxical. Thus, linguistic evidence would suggest that whereas belief (for example) is required for knowledge, absolute certainty is not.

Even if one denies the force of the linguistic evidence for fallibilism (perhaps holding that ordinary linguistic practice is just mistaken), there is more to be said in its defense. For it is open to the fallibilist to concede the concept of knowledge to the skeptic, and invent a new concept—call it knowledge*—which, in lieu of absolute certainty, requires only good reasons, or justified beliefs, or reliable processes, or some other preferred criterion. In fact, it is not only open to the fallibilist to do this, but it seems *necessary* to do this; for we need a way to distinguish between beliefs that are the result of lucky guesses and beliefs that are arrived at through good reasoning (or reliable processes, or what have you). This inquiry into knowledge*—call it epistemology*—would clearly be a rich and fruitful enterprise; after all, it

is arguably what most epistemologists have been doing for centuries anyway. In contrast, the pursuit of, and inquiry into, knowledge (in terms of absolute certainty) would be rather impoverished, as it would be relevant in very few (if any) cases. In this imagined scenario, it would be much more desirable and fruitful to study and think about knowledge*, rather than knowledge. Thus, it seems best to avoid inventing a new concept and instead to simply drop the absolute certainty criterion from the old concept. This is the practical consideration in favor of fallibilism (cf. again Feldman 2003, 123).

So why not endorse fallibilism? Well, some have argued that endorsing fallibilism is *madness*. For example, David Lewis says the following (1996, 691): "To speak of fallible knowledge, of knowledge despite uneliminated possibilities of error, just *sounds* contradictory." He later laments (1996, 692) that "We are caught between the rock of fallibilism and whirlpool of skepticism. Both are mad!" He then supports this contention with, among other reasons, the following plea:

> If you are a contented fallibilist, I implore you to be honest, be naive, hear it afresh. "He knows, yet he has not eliminated all possibilities of error." Even if you've numbed your ears, doesn't this overt, explicit fallibilism *still* sound wrong? (Lewis 1996, 692)

Given the choice between fallibilism and skepticism, Lewis admits that he would choose fallibilism;[44] but only as the lesser of two madnesses. This is hardly a ringing endorsement.

Fantl and McGrath (2009) make a similar point. They agree that fallibilism is mad, in part because the fallibilist is forced to affirm these "clashing conjunctions" (2009, 15):

(8) I know that *p* but there is a chance that not-*p*.

(9) I know that *p* but it's possible that not-*p*.

[44] To draw another parallel between the consequence argument and the skeptical argument, Lewis's admission here is reminiscent of van Inwagen's (1983, 149–150) admission that he must choose between "the puzzling" (roughly, the compatibility of freedom and indeterminism) and "the inconceivable" (the compatibility of freedom and determinism). He chooses the puzzling. (Thanks to John Fischer for this point.)

They then ask:

> Don't these just sound wrong, *at least when one is careful to read both conjuncts as simultaneously endorsed,* rather than reading the second conjunct as a correction to the first? ... If these sorts of statements are often true, as the fallibilist must admit, why the discomfort? If they're right, why do they feel so wrong? (2009, 15, emphasis mine)

While I agree that conjunctions such as (8) and (9) do have somewhat of a "clashy" sound to them, the accusation of madness rings a bit hollow to me. Nevertheless, the problems with fallibilism seem to be significant enough to justify looking elsewhere for a way to avoid the absolutist notion of knowledge that leads to skepticism.

4.5 Sensitivity

An alternative strategy (for rejecting the absolutist notion of knowledge) begins by noting that there is in fact an account of knowledge (or rather, a family of related accounts of knowledge) that allows us to make a distinction, at least indirectly, between possibilities that need to be ruled out and those that don't. I will call these accounts of knowledge *sensitivity*-based accounts, since they make sensitivity (of the relevant belief) a necessary condition on knowledge.

So what is it for a belief to be sensitive? According to Nozick (1981), who was one of the first to explicitly require sensitivity for knowledge (but see also Dretske 1971),[45] *S* knows that *p* only if

(10) *S* wouldn't believe that *p,* were *p* false.

[45] Nozick (1981) also added what might be called an "adherence" condition to his analysis of knowledge: If *p* were true, then *S* would believe that *p*. A belief that satisfies both sensitivity and adherence "tracks the truth." I will set aside the adherence condition, in part to focus on what Dretske and Nozick share in common (namely the sensitivity condition), and also because the sensitivity condition is what has gotten the lion's share of the focus in the subsequent literature. In addition, the sensitivity condition is what provides the pressure to deny closure, and is thus more relevant to the discussion later in the chapter.

According to the standard evaluation of subjunctive conditionals such as (10), we examine the nearest possible world in which *p* is false and ask whether *S* believes that *p* in that world. If she does, then her belief is not sensitive; hence she doesn't know that *p*. But if she doesn't believe that *p* in that world, then her belief is sensitive. In other words, *S*'s belief that *p* is sensitive just in case in the nearest possible world in which *p* is false, *S* does not believe that *p*.

One benefit of the sensitivity condition is that it allows us to solve the Gettier (1963) problem that afflicts the traditional account of knowledge as justified true belief. Sensitivity theorists solve this problem by substituting the sensitivity condition for the justification condition.[46] Consider, for example, one of Gettier's classic cases: Smith believes, with justification, that Jones owns a Ford; on this basis he justifiedly infers that

(11) Either Jones owns a Ford, or Brown is in Barcelona.

Smith has no reason to believe that Brown is in Barcelona, but it turns out that he is. Moreover, Smith's belief in Jones's car ownership, though justified, turns out to be false. Hence, Smith has a true, justified belief in (11), and yet he doesn't *know* that (11) is true. So the proponent of the traditional JTB analysis of knowledge gets it wrong with respect to (11). But notice what happens when we replace the justification condition with a sensitivity condition. When we apply the sensitivity condition, we find that, in the nearest possible world in which (11) is false (i.e., the nearest possible world in which Brown is not in Barcelona), Smith still believes that it's true—since, in that world, he still infers (11) on the basis of his (faulty) evidence regarding the first disjunct. And since Smith would continue to believe that (11) is true even if it were false, his belief that (11) is not sensitive, and hence does not constitute knowledge.

[46] An alternative suggestion would be that sensitivity should supplement, rather than replace, the justification condition.

Apropos of current purposes, the sensitivity condition also provides us with a principled distinction between the possibilities we need to rule out and the possibilities we don't need to rule out in order to have knowledge. When I'm asking myself whether I know various ordinary propositions, such as the proposition that I had pancakes for breakfast this morning, I first ask whether I would believe it in the nearest possible world in which it's false. And it seems that I would not: the nearest possible world in which I don't eat pancakes for breakfast is a world where I decide to eat something else for breakfast. And in this alternative world, I wouldn't believe that I had pancakes for breakfast. Thus, my belief that I had pancakes for breakfast is sensitive.

It appears, then, that sensitivity does provide a principled way of drawing the line we need: the line gets drawn at the nearest world in which the belief in question is false. Worlds that are more remote than that (i.e., worlds that differ from the actual world to a greater degree) need not be considered. Unfortunately (from the perspective of those who want to dismiss the skeptical argument because it begs the question), there are some compelling reasons that speak against adopting a sensitivity-based account of knowledge.

4.6 Problems with sensitivity-based accounts

Perhaps the most serious problem with sensitivity-based accounts of knowledge is that the sensitivity requirement forces us to abandon the closure principle in (1): If *S* knows that *p,* and knows that *p* entails *q,* then *S* knows that *q.* The easiest way to see this is to examine the following Moorean argument:

(12) I know that I had pancakes for breakfast this morning.

(13) I know that if I had pancakes for breakfast this morning, then I'm not plugged into the Matrix.

Therefore, by (1),

(14) I know that I'm not plugged into the Matrix.

The problem is that whereas both (12) and (13) meet the sensitivity condition, (14) does not. The belief in (12) meets the sensitivity condition because, as we saw above, my belief that I had pancakes for breakfast is sensitive. The belief in (13) is also sensitive, albeit trivially so. (There's no possible world in which I had pancakes for breakfast this morning and yet was plugged into the Matrix [given the details of how the Matrix works], so there's no possible world in which it's false that if I had pancakes for breakfast, then I'm not plugged into the Matrix. The sensitivity condition is trivially met because the relevant subjunctive conditional is vacuously true.) But what about (14)? In order for my belief that I'm not plugged into the Matrix to be sensitive, we have to examine the nearest possible world in which it's false that I'm not plugged into the Matrix (i.e., the nearest possible world in which I'm plugged in). In this world, even though it's false that I'm not plugged in, I would still believe that I'm not plugged in. Therefore, my belief that I'm not plugged into the Matrix is *not* sensitive and (14) is false. The only option here, at least for an advocate of the sensitivity requirement, is to reject the closure principle in (1).

To recap: sensitivity-based accounts allow us to reject the absolutist notion of knowledge that gives the skeptical argument traction. But the sensitivity requirement is problematic insofar as we want to maintain our intuitive commitment to epistemic closure. (Later in the chapter we'll explore the question of whether rejection of closure might be a viable strategy.) One natural move, in light of this conundrum, is to abandon sensitivity and examine the viability of some other modal conditions on knowledge—in particular, the *safety* condition.

4.7 Safety

Safety-based accounts of knowledge, as popularized by Sosa (1999, 2000, 2002), Williamson (2000), and Pritchard (2002, 2005), are so-called because they make a belief's being *safe* a

necessary condition of that belief's counting as knowledge. According to such accounts, *S* knows that *p* only if

(15) If S were to believe that *p,* then *p* would be true.

What (15) tells us is that when we are examining the status of *S*'s belief that *p,* we need to examine a range of close possible worlds in which *S* believes that *p* and ask whether *p* is true in all (or nearly all) of those worlds.[47] If the answer is yes, then the belief is safe, and—assuming that it's also true in the actual world—constitutes knowledge. We've already seen that my belief that I had pancakes for breakfast is sensitive; but is it also safe? Well, is it true in all of the nearby worlds in which I believe it? It seems that it is, for it's not very often that, over the course of a day, I forget what I had for breakfast that morning. Given the rarity of the situation in which I'm mistaken about what I had for breakfast, it seems plausible to suppose that the nearest world in which I falsely believe that I had pancakes for breakfast is quite remote from the actual world. Thus, I can (and do) know that I had pancakes for breakfast this morning.

It appears, then, that the safety requirement also provides a principled way of drawing the line we need. We need to look at all of the nearby worlds in which the agent in question believes the proposition in question, but we do not need to look any farther than that. There is, of course, the thorny issue of how to decide which worlds count as "nearby." That would need to be worked out, were we considering a safety-based account for adoption. For our purposes, however, we are only using the safety account to establish the possibility of distinguishing between epistemically different classes of scenarios (i.e., those which preclude knowledge and those which don't). So as long as there's at least some hope for a

[47] Opinions differ on whether the safety condition applies to all or just nearly all nearby worlds. Although I'm inclined to follow Pritchard (2005) in endorsing the stronger condition, little here will depend on that so I'll remain officially neutral.

criterion of nearness (and I see no reason to think that such a criterion is hopeless), the safety condition seems capable of doing what we want it to do.

At this point we've seen that both sensitivity and safety provide us with the principled distinction that we've been looking for. In the case of sensitivity, however, this distinction appears to come at a cost—namely that of denying closure.[48] Does safety come with a similar cost?

4.8 Can safety preserve closure?

Since at this point we are considering only the question of whether skepticism begs the question, and not the question of which analysis of knowledge is the best, our vetting of the safety condition can be less thorough than it might otherwise need to be. We can focus on two questions. First: Does safety, like sensitivity, allow us to solve the Gettier problem? And second: Does safety, unlike sensitivity, allow us to preserve closure?

As we saw above, one benefit of sensitivity is that incorporating it into an analysis of knowledge immunizes that account from the Gettier problem—so we want to make sure that we're not losing that benefit by adopting a safety condition. As it turns out, safety does preserve this benefit. In the Gettier case as described above, the relevant belief is Smith's belief that

(11) Either Jones owns a Ford, or Brown is in Barcelona.

This belief that (11) is safe if and only if (11) would be true, were Smith to believe it. But there are plenty of nearby worlds in which Smith continues to believe (11) and yet (11) is false. (This is because we can change various minor details about the scenario without changing the fact

[48] Not all sensitivity theorists will agree that a sensitivity-based account is forced to deny closure. (See, for example, DeRose (1995), who thinks that a denial of closure leads to an "abominable conjunction"—of an ordinary knowledge attribution on the one hand, and a denial of knowledge that a skeptical hypothesis fails to obtain on the other hand.) But describing and evaluating the theoretical machinery required for maintaining both sensitivity and closure would take us too far afield from present purposes.

that Smith believes that (11).) Therefore Smith's belief that (11) is not safe—and hence not knowledge. The safety theorist can avoid the Gettier problem.

But what about the closure principle? At first glance, it seems that the proponent of safety is in just as much trouble as the proponent of sensitivity. For suppose now that Smith, having gotten past the troublesome business about Jones and Brown, happens to be driving through fake barn country. (In fake barn country, real barns are painted red while fake barns are painted blue.[49]) Smith sees a red barn. Now consider the question of whether Smith knows that he's looking at a barn. According to the safety theorist, Smith doesn't know that he's looking at a barn, because the belief that he's looking at a barn, though true, isn't safe. After all, there are plenty of nearby worlds in which Smith looks at a blue barn façade, and hence believes that he is looking at a barn, and yet is mistaken. So it's not the case that Smith is looking at a barn in all nearby worlds in which he believes that he's looking at a barn—which means that his belief that he's looking at a barn is not safe and hence not knowledge. But this verdict presents a problem for the safety theorist, because the following argument—which relies on closure—appears to establish that Smith *does* know that he's looking at a barn:

(16) Smith knows that he's looking at a red barn.

(17) Smith knows that if he's looking at a red barn, then he's looking at a barn.

So, (18) Smith knows that he's looking at a barn.

Closure guarantees that (18) is true, given that (16) and (17) are true—but, again, according to a straightforward safety account, (16) and (17) are true while (18) is false.

Are there ways to tweak the safety account such that it can withstand this criticism? According to Sosa (2004, 292–94), the answer is yes. Sosa attempts to preserve closure by pointing out that there are two reasons why (16) can't be used in an argument for (18). The

[49] This is a variation on the fake barn case considered by Pritchard (2005, 167), who attributes it to Kripke via Kvanvig (2004).

first reason is simply that it's false. According to Sosa, Smith's belief in (16) is inferential, and the inference is based on at least one belief that isn't safe (and hence cannot be known). Moreover, an inferential belief is safe only if the beliefs from which it is inferred are themselves safe. To be more specific: When Smith looks at the red barn, he believes that he's looking at something red, and he believes that he's looking at a barn. He then infers that he's looking at a red barn. Here is the inference from Smith's perspective:

(19) I'm looking at something red.

(20) I'm looking at a barn.

So, (21) I'm looking at a red barn.

For Smith's belief in (21) to be safe, and thus count as knowledge (assuming, as we are, that it's true), it must be inferred from premises that are also safely believed. Are (19) and (20) both safely believed? In the scenario as described, Smith's belief in (19) is safe: there are no shenanigans involving red things, so in all nearby worlds in which Smith believes that he's looking at something red, that thing is indeed red. But his belief in (20) is not similarly safe; given that Smiths finds himself in fake barn country, there are numerous scenarios in which he believes that he's looking at a barn, and yet it's false that he's looking at a barn (because he's looking at a barn façade). (This, recall, is what was supposed to cause trouble for the safety theorist in the first place.) Thus, Smith's belief that he's looking at a red barn (i.e., the belief in (21)) is not safe and he can't know that what he's looking at is a red barn. And since Smith can't know (21), it follows that we must reject (16). In other words: the problematic argument never even gets going, on account of (16) being false.

The second reason why (16) can't be used in an argument for (18) should now be evident as well. If Sosa is right, then Smith's putative knowledge that he's looking at a red barn *relies* on, rather than establishes, his knowledge that he's looking at a barn. So the argument for (18) above may indeed be problematic—but not in any way that causes trouble for the safety theorist.

Although I find this argument convincing, it's worth noting that closure is constantly threatening to endanger these otherwise plausible (although perhaps not ultimately successful) proposed conditions on knowledge. Given this crucial role that the epistemic closure principle plays in some of the discussions surrounding the skeptical argument, it's worth revisiting the question of whether there are good reasons to simply reject it.

4.9 Can we reject the epistemic closure principle?

As we consider the viability of rejecting epistemic closure, let us begin with Michael Slote's (1982) dismissal of the principle. His claim is that closure fails because people don't always make all of the inferences they can:

> It is generally agreed that '*A* knows that *p*' and '*A* knows that (*p* ⊃ *q*)' do not entail '*A* knows that *q*' for appropriate substituends. People may fail to make inferences they are entitled to make. (Slote 1982, 11)

But as Fischer (1995) points out, this move is a bit quick. Is it really so obvious that the epistemic version fails? What reasons are there for thinking that it does? Fischer's first point is that this fact about us—that we may fail to make inferences that we're entitled to make—may be sufficient to undermine the principle that knowledge is closed under (mere) entailment, but it is *not* sufficient to undermine the principle that knowledge is closed under *known* entailment. If *S* knows that *p,* and if *p* entails *q,* then it is easy to see how *S* might fail to know that *q*: *S* might fail to know that *p* entails *q* and thus fail to make the relevant inference. But if *S* does in fact know that *p* entails *q,* then it's not so clear that *S* might fail to know that *q*. In any case, *S* is certainly *in a position* to know that *q,* and it is certainly not "generally agreed" that *S* doesn't know that *q*.

Another strategy for denying epistemic closure begins with the premise that knowledge is closed under (known) entailment only if each necessary condition for knowledge is also closed under known entailment (Brueckner 1985, 91). If this premise is correct and if, for example, sensitivity is required for knowledge but fails to be closed under

known entailment (as appears to be the case), then the epistemic closure principle is invalid. (Brueckner goes on to argue that there are in fact no necessary conditions on knowledge that clearly fail to be closed under known entailment.) But Warfield (2004, 38) has argued that the initial premise is false in its general form (where R is some closure relation): "The failure of some necessary condition on knowledge to have some property (example: closure under R) does not imply that knowledge itself does not have the property." He points out that knowledge could have the property in virtue of some other necessary condition having the property, or it could have the property in virtue of some interaction between necessary conditions. Thus it seems that this second strategy is not ultimately very promising.

The third and final strategy that we will consider is based on some interesting and important examples from Dretske (1970), Vogel (1990), and Hawthorne (2004). But before we take a look at the examples, let's make explicit what seems to me the most intuitive and powerful motivation for denying closure—namely that if we're forced to accept it, then we're also forced to admit that we don't know a lot of things that we intuitively and plausibly think we know. Recall the final steps of the skeptical argument:

(4) If I know that I had pancakes for breakfast this morning, then I know that I'm not plugged into the Matrix.

(5) But I don't know that I'm not plugged into the Matrix.

So, (6) I don't know that I had pancakes for breakfast this morning.

For someone who, like most of us, wants to reject the conclusion in (6), the weakest point of the argument appears to be (4), which is entailed by the closure principle (1). If we can impugn (4), then we can reject (1). The Moorean move, in which we insist on the falsity of (6)—and thus the falsity of (5) via (4)—is always an option, but it's difficult to argue for the legitimacy of such a move. This is of course consistent with it nevertheless *being* a legitimate move, but it still strikes me as something of a last resort. Would it be better to reject closure

instead? Can we muster a good argument for rejecting closure? Next we will examine what I take to be the strongest reasons for such a rejection.

An important feature of the skeptical argument, as we saw above, is the fanciful nature of the skeptical hypothesis. It seems to us that the possible world in which the Matrix scenario obtains is quite remote from the actual world. (Although of course the status of this intuition is one of the things that the skeptical argument is designed to challenge.) And if rejecting closure requires taking seriously fanciful hypotheses such as the Matrix scenario, then we might be tempted to accept closure and instead focus on rejecting the relevance of the skeptical hypothesis.[50] For it seems that the more far-fetched the scenario, the easier it is to deny the hypothesis that the scenario obtains.

But matters are not quite so simple. For if I'm right in thinking that fanciful hypotheses are easier to reject than closure, then we can make trouble for those who want to affirm closure by describing ordinary, unremarkable scenarios and then plugging in those scenarios as hypotheses for the skeptical argument. If affirming closure leads us to deny that we're in a fanciful skeptical scenario, then that's a cost—but perhaps not a significant cost. If, however, affirming closure leads us to deny that we're in an ordinary scenario, then the cost seems more significant. As the hypothesis becomes more mundane, the cost of denying it (on the basis of closure) becomes greater. One strategy, then, is to put pressure on proponents of closure by looking for skeptical arguments featuring mundane hypotheses. As it turns out, such arguments have recently proliferated.[51] The first argument involves the unlikely but not entirely fanciful hypothesis that someone has won the lottery. Here's what it looks like if we plug that hypothesis into the skeptical argument:

> Suppose someone of modest means announces that he knows he will not have enough money to go on an African safari this year. We are inclined to treat such a judgment as true, notwithstanding various far-fetched possibilities in which that person suddenly acquires a great deal of money.

[50] Or perhaps, following Williamson (2000), we could somehow reject the claim that the skeptical scenario is indistinguishable from (what we take to be) the actual scenario.

[51] See especially Hawthorne (2004), who draws inspiration from Vogel (1990).

> We are at some level aware that people of modest means buy lottery tickets from time to time, and very occasionally win. And we are aware that there have been occasions when a person of modest means suddenly inherits a great deal of money from a relative from whom he had no reason to expect a large inheritance. But despite all this, many normal people of modest means will be willing, under normal circumstances, to judge that they know that they will not have enough money to go on an African safari in the near future. And under normal circumstances, their conversational partners will be willing to accept that judgment as correct.
>
> However, were that person to announce that he knew that he would not win a major prize in a lottery this year, we would be far less inclined to accept his judgment as true. We do not suppose that people know in advance of a lottery drawing whether they will win or lose. But what is going on here? The proposition that the person will not have enough money to go on an African safari this year entails that he will not win a major prize in a lottery. If the person knows the former, then isn't he at least in a position to know the latter by performing a simple deduction? (Hawthorne 2004, 1–2)[52]

That "simple deduction," of course, is licensed by the closure principle. And it seems that insofar as we feel uncomfortable claiming to know that we won't win the lottery (assuming the purchase of a ticket), we should feel uncomfortable endorsing closure.

The same discomfort arises with respect to various non-observational beliefs. I think I know that Barack Obama is the President of the United States. But if I know this, then I can perform a simple deduction and come to know that he hasn't died in a tragic plane crash this morning. But do I really know that Obama hasn't died in a plane crash this morning? Intuitively, it seems that I do not know this.[53] Here we have an example of what we might call a skeptical hypothesis: that President Obama has died in a plane crash this morning. Most of us possess the dispositional, non-observational belief that this hypothesis is false. But when we are pressed on this belief that Obama has not recently died in a plane crash, we might be reluctant to claim that it is knowledge. Closure, however, forces the claim that this *is* knowledge, and as a result we might be further tempted to reject closure.

The same puzzle arises for beliefs about the future—as evidenced by this additional example from Hawthorne (2004), who is himself drawing from Vogel (1990). Consider the

[52] The footnotes have been deleted from this passage.

[53] This particular example is made even more compelling by the April 2010 plane crash (See Stack 2010) that killed Polish President President Lech Kaczynski (along with many other members of the Polish government).

(slightly more mundane [and frankly somewhat disturbing]) hypothesis that I'll have a heart attack sometime in the next six months. It seems that I know that I will be in San Diego in April, because there is a conference there that I plan to attend. But my being in San Diego in April entails that I don't have a heart attack in the next six months. Thus, if I know that I will be in San Diego in April, then I can perform another simple deduction and come to know that I won't have a heart attack in the next six months. Heart attacks unfortunately happen more often than lottery winnings,[54] which means that closure has apparently forced us to claim knowledge of the falsity of an even more mundane hypothesis.

We can continue to move toward the mundane end of the skeptical hypothesis continuum, and consider the hypothesis that my car was recently stolen (cf. again Vogel 1990). It's more likely that my car will be stolen than that I'll have a heart attack,[55] so let's consider the proposition that my car is parked in Lot F (where I left it this morning). It seems that I know that my car is parked in that lot. But that of course entails that it hasn't been stolen since I parked it this morning.[56] Thus, if I know that my car is parked in Lot F then I can perform a simple deduction and come to know that my car hasn't been stolen today. But I have a strong intuition, which I suspect is widely shared, that right now I don't in fact know that my car hasn't been stolen. (In order to know, wouldn't I have to go and check?) If closure forces us to the conclusion that we *do* know such a thing, then, again, we might want to think seriously about rejecting closure.

---

[54] Although I don't have exact numbers for the odds of dying of a heart attack, the annual risk of dying from heart *disease* is 1 in 397 (according to the Harvard Center for Risk Analysis, from U.S. Centers for Disease Control and Prevention data, as cited in Ciulla et al. (eds.) 2007). I think it's safe to extrapolate from these numbers and claim that I am more likely to suffer from a heart attack than win the lottery, the odds of which are usually 1 in millions.

[55] Whether this is true will depend on the locale, but the annual risk of having one's car stolen can be as high as roughly 1 in 130 (according to the National Insurance Crime Bureau).

[56] Strictly speaking, of course, its being parked in Lot F only entails that it isn't *currently* stolen; it could have been stolen and returned since I parked this morning.

4.10 The cost(s) of denying closure

Earlier we examined two putative requirements for knowledge, sensitivity and safety, that appear to be in tension with closure (although if Sosa is right, there's a way to be a safety theorist while holding on to closure). We have also seen some powerful examples that tempt us to reject closure. Before we do so, however, we should consider the costs.

Recall that the benefit of denying closure is that it allows us to claim to know things we think we know without forcing us to affirm that we know things we *don't* think we know. (It also frees us up to adopt, e.g., a sensitivity-based account of knowledge.) In other words, denying closure allows us to deny (4):

> (4) If I know that I had pancakes for breakfast this morning, then I know that I'm not plugged into the Matrix.

But this benefit not only comes with a cost; it *is* a cost. The problem with denying (4)—the first cost of denying closure—is simply that it's incredibly counterintuitive. If I know that I had pancakes for breakfast, and if I know that having pancakes for breakfast entails not being plugged into the Matrix, then how can I fail to know that I'm not plugged into the Matrix? But not all people will share this intuition. As such, it will be helpful to examine some additional and more specific costs of rejecting closure.

The second cost of denying closure is that, as Hawthorne (2004, 39) puts it, such a denial "interacts disastrously with the thesis that knowledge is the norm of assertion." The knowledge norm of assertion enjoins us to assert only what we know. Although it's an open question whether such a norm really does apply to assertion, it does seem at the very least to be a useful guide to when we should and should not assert. But if I reject closure, while endorsing the knowledge norm, and allow that I don't know that skeptical hypotheses fail to obtain, then I find myself in the following predicament: Someone asks me if I had pancakes for breakfast this morning; I reply in the affirmative (i.e., I assert that I had pancakes this morning). He then asks me if I would agree that if I had pancakes for breakfast, then I'm not

plugged into the Matrix; I again reply in the affirmative. Finally, he asks me if I would agree that I'm not plugged into the Matrix. To this last question I must answer in the negative. In this imagined situation, I am in the unenviable position of affirming the premises of a modus ponens while rejecting the conclusion.

One of the ways to construe this point is as an argument against a sensitivity-based account of knowledge. If one is led by the sensitivity condition to reject closure, then one is also forced to reject the knowledge norm of assertion. Given this situation—and the *prima facie* plausibility of both closure and the knowledge norm of assertion—it might seem preferable to reject the sensitivity account of knowledge instead. And in fact there are additional reasons to reject the sensitivity account. Consider a case from Vogel:

> *Hole-In-One Case*. Sixty golfers are entered in the Wealth and Privilege Invitational Tournament. The course has a short but difficult hole, known as the "Heartbreaker." Before the round begins, you think to yourself that, surely, not all sixty players will get a hole-in-one on the "Heartbreaker." (Vogel 1999, 165)[57]

This belief (that not all players will ace the Heartbreaker) seems to count as knowledge; but is it sensitive? It seems not. For in the nearest possible world in which all sixty players ace the Heartbreaker, I would still believe, falsely, that not all of them will do so. There is a possible world in which holes-in-one are as commonplace as pars are in the actual world, and in which all sixty golfers do get a hole-in-one, and perhaps in *that* world I would not believe that not all sixty will get a hole-in-one. But surely that world is not as similar to the actual world as the world in which holes-in-one remain a rarity, and yet all sixty golfers happen to get one on the Heartbreaker. Thus, given the choice between closure and sensitivity (as a requirement for knowledge), we might prefer to give up on sensitivity rather than closure.[58]

---

[57] Although Vogel introduces this case to refute a *salience* criterion on knowledge, it applies to the sensitivity condition as well.

[58] Examples such as this also seem to recommend safety over sensitivity, at least as a requirement on knowledge.

A third cost of rejecting closure, and the final one that we'll consider, also comes from Hawthorne (2004, 39–41). This cost is that if we reject the closure of knowledge under known implication, then we will also have to reject other principles that are just as intuitive—if not more intuitive—than knowledge closure. For example, consider a version of what Hawthorne calls the *equivalence principle:*

(22) If I know that *p,* and if I know *a priori* that *p* and *q* are equivalent, then I know that *q*.

This principle seems quite plausible, as does the *distribution principle:*

(23) If I know that *p* and *q,* then I know that *p* and I know that *q*.

But if we are led (by sensitivity concerns, or a discomfort with fallibilism, or some other consideration) to deny

(14) I know that I'm not plugged into the Matrix

then we again run into problems. For given the truth of

(12) I know that I had pancakes for breakfast this morning,

we can use (22) to deliver the result that

(24) I know that I had pancakes for breakfast this morning and I'm not plugged into the Matrix.

But applying (23) to (24) gives us (14)—which, according to all but the staunchest Mooreans and fallibilists, is what we're supposed to be denying. So those who want to reject closure

must also reject either the equivalence principle (22) or the distribution principle (23)—neither of which is an appealing prospect.

Thus what we have seen is that even though there are some compelling examples that might lead us to reject the closure of knowledge under known implication, there are also significant costs to such a rejection. In particular, abandoning closure forces us to do violence to several intuitive epistemic judgments and principles. In light of these considerations, and despite the lack of a completely satisfactory response to the lottery paradox, I propose that we maintain a provisional commitment to the principle that knowledge is closed under known implication. Future argumentation might lead us to adjust the balance sheet, but for now it seems that the costs of rejecting closure outweigh the benefits.

4.11 Conclusion

In this chapter we have taken a closer look at the skeptical argument in epistemology. As a way of getting clearer on the argument, and some of the preemptive responses to it, we considered the question of whether the skeptical argument begs the question. We saw that if the concept of knowledge that the skeptic is operating with implies that we must rule out *all* possibilities, then the non-skeptic might be inclined to accuse the skeptic of begging the question. But the issue is tricky. On the one hand, it's difficult to come up with independent reasons for endorsing the absolutist notion of knowledge that gives the skeptical argument traction. On the other hand, it's also difficult to come up with a principled way of distinguishing between possibilities that preclude knowledge and those that don't. We looked at two putative requirements on knowledge—the sensitivity condition and the safety condition—which produce the desired division between possibilities that need to be ruled out and possibilities that don't, but which also conflict with the epistemic closure principle. (Sensitivity-based accounts of knowledge are particularly problematic in this regard.) We also saw that closure threatens to endanger much of our ordinary knowledge of everyday propositions—even if we restrict ourselves to considering hypotheses involving relatively

mundane occurrences, such as heart attacks and car theft. Because of these conflicts, we considered the costs and benefits of giving up on the closure principle. My own view is that the costs of rejecting closure outweigh the benefits, and as a result we should accept the closure principle if at all possible.

The upshot of all this is that the preemption responses to the skeptical argument (i.e., rejecting it on the grounds that it begs the question, or rejecting the closure principle) do not appear to boast much promise. It seems, then, that the best way to proceed is to consider the argument on its own terms—including both the absolutist notion of knowledge and the epistemic closure principle. If we can respond to the skeptical argument within these constraints, then we will realize two benefits. First, we will have a stronger response to the skeptic, in virtue of already having granted him two key claims. Second, we will have an explanatory schema that will serve us well as we respond to the consequence argument. Or so I shall argue.

• Chapter 5 •

The contextualist gambit

5.1 Introduction

We saw in the previous chapter how the fallibilist response to the skeptical argument in epistemology is less than fully satisfying. And we saw that an alternative response, which denies closure, also fails to be fully satisfying. If we want to import insights from epistemology to help us formulate a new response to the consequence argument, then we need to look further. I think there are two pieces that will be required as we put together this new response, and in this chapter I will examine and develop one of those pieces: *contextualism*.

Recall from Chapter 4 David Lewis's lament about what appears to be a choice between the rock (of fallibilism) and the whirlpool (of skepticism). He resolves to dodge the choice:

> Better fallibilism than skepticism; but it would be better still to dodge the choice. I think we can. We will be alarmingly close to the rock, and also alarmingly close to the whirlpool, but if we steer with care, we can—just barely—escape them both. (Lewis 1996, 692)

Although I will remain officially neutral on the question of whether Lewis's contextualist proposal (or subsequent proposals) succeeds in disarming the skeptic, I will be arguing that the metalinguistic mechanism he identifies can be incorporated into a response to the consequence argument. Thus, in order to get clearer on what exactly the contextualist move is, I will be examining Lewis's view (1979a, 1996) for the better part of this chapter.

5.2 Lewis's contextualism: scorekeeping in a language game

Lewis's contextualism begins to take shape with his "Scorekeeping in a Language Game" (1979a), in which he explores some of the ways in which conversational rules of accommodation work. If we view a conversation as analogous to a game (e.g., a baseball game), then at any time during the conversation we can specify the "score" of the conversation in terms of the previous behavior of the "players" (i.e., the individuals involved in the conversation), as well as the various rule-governed ways in which the score can change (what Lewis calls the "kinematics of score"). Having specified the kinematics of score, we can then define "correct play" in terms of the relation between the players' current behavior and the score. A correct play in a conversation occurs when, to put it roughly, one of the players utters a sentence that is true, or perhaps acceptable in some other way. Moreover, conversations are often regulated by directives (e.g., of cooperation or conflict) that dictate a certain conversational direction.

On this picture, conversations are also governed by *rules of accommodation*—governed in the sense that the evolution of a conversational score tends (although not inexorably) toward making whatever occurs count as a correct play. Thus, if a particular statement is made that requires a certain presupposition, then, within certain limits, that presupposition is thereby created and added to the context of the conversation. Similarly, if a conversational statement requires that some definite description "the F" have a certain denotation, then typically that description will have that denotation. Lewis (1979a, 348) provides the following example: "The pig is grunting, but the pig with the floppy ears is not grunting." For this statement to be true, the second instance of "the pig" must denote a different pig than the one denoted by the first instance of the description. These two examples (of presupposition and definite description) involve two different rules of accommodation (which govern two different components of the conversational score), but Lewis lays out the general scheme as follows:

> If at time t something is said that requires component s_n of conversational score to have a value in the range r if what is said is to be true, or otherwise

> acceptable; and if s_n does not have a value in the range r just before t; and if such-and-such further conditions hold; then at t the score-component s_n takes some value in the range r." (Lewis 1979a, 347)

The most relevant example of this phenomenon, at least for present purposes, is *vagueness* (Lewis 1979a, 352–54). To say, for example, that the sentence "Fred is tall" (where Fred is a borderline case of tallness) is vague is to say that the sentence is true with respect to some reasonable ways of drawing the line between tallness and shortness, and false with respect to other reasonable ways of drawing the line. Where exactly that line is drawn is going to be a component of conversational score. Suppose that we're having a conversation about Fred's height, and I say the following: "No, Fred's not tall; in fact, he didn't make the volleyball team in college because he was too short!" Although Fred is, by hypothesis, a borderline case of tallness, the truth (or acceptability) of my claim requires that the standards of tallness be fairly strict. And so, according to the rule of accommodation, those standards shift and that particular component of our conversational score now has a different value.

According to Lewis, if a sentence is true over a large enough range of these delineations, then it is *true enough*—in which case we are typically willing to assert it, assent to it, accept it, and so forth. Moreover, what determines whether a sentence is true enough are the *standards of precision* that are in play. These standards of precision (which, perhaps unfortunately, are themselves vague) can be viewed as another component of conversational score, and hence governed by a rule of accommodation. But the key point here is that whereas the relevant standards of precision can be raised or lowered over the course of a conversation, it is easier to raise the standards than it is to lower them. And even when the standards have been lowered, there remains a residue of unacceptability. Thus Lewis:

> If the standards have been high, and something is said that is true enough only under lowered standards, and nobody objects, then indeed the standards are shifted down. But what is said, although true enough under the lowered standards, may still seem imperfectly acceptable. Raising of standards, on the other hand, manages to seem commendable even when we know that it interferes with our conversational purpose. Because of this asymmetry, a player of language games who is so inclined may get away with it if he tries to

> raise the standards of precision as high as possible—so high, [for example], that no material object whatever is hexagonal. (Lewis 1979a, 352–53)

If Lewis's picture is accurate, then we can say the following. Within a particular conversational context, once the standards of precision are raised, no statement that can only be true enough relative to lower standards of precision will ever again seem fully acceptable.

Lewis then applies this insight to Peter Unger's (1979) argument for skepticism. Unger's argument turns on the notion that "certainty" is a member of the class of *absolute terms*. And one feature of absolute terms is that they are rarely, if ever, appropriately predicated of a person or object. For example, "flat" is also supposed to be an absolute term because for almost anything that we're typically willing to call flat, there's something else we can think of that is flatter than what we originally called flat. I wouldn't hesitate to describe my driveway as flat, but of course there are countless surfaces that are flatter than my driveway. And since nothing can be flatter than something that's truly flat, when I'm reminded of these other surfaces (e.g., a glass table) it seems that I'm forced to acknowledge that I was mistaken in referring to the driveway as flat. And Unger makes a similar move with respect to certainty, arguing that hardly anyone is ever certain of anything.

Lewis's diagnosis of the situation here is that Unger's arguments are trading on the relevant rules of accommodation. In order for the claim that my glass table is flatter than my driveway to be true, the standards of precision (for what counts as flat) need to be raised such that whatever bumps there may be on the surface of my driveway are now relevant to the question of whether it's flat (whereas before they were *not* relevant). And, because the rule of accommodation is at work, the standards are indeed raised when that claim is made. But the crucial point (the first crucial point) is that these stricter standards of precision do not apply to, and thus do not render unacceptable, previous statements made in previous contexts (when looser standards were in effect). As a result, there is no contradiction between "My glass table is flatter than my driveway," which is true enough according to raised standards of precision, and "My driveway is flat," which was true enough according to

the original (unraised) standards of precision—"any more than 'It is morning' said in the morning contradicts 'It is afternoon' said in the afternoon" (Lewis 1979a, 353).

The second crucial point is that the context in which standards have been raised is not in any way superior or preferable to the context in which standards remain unraised.[59] In fact, a context in which, for example, a driveway does not count as flat might even be describable as an *unusual* context (cf. Lewis 1979a, 353). And certainly a context in which a glass table doesn't count as flat comes across as an unusual context.

The final example of a rule of accommodation that we'll consider is the rule that governs *relative modality*. As Lewis points out (1979a, 354–55), ordinary language modal verbs—e.g., "can," "must," and "knows"—do not typically express absolute logical or metaphysical possibility. Instead, these verbs express relative modalities, which is to say that various possibilities can be ignored when we are evaluating statements in which they occur. Thus the boundary between possibilities that are relevant and possibilities that can be ignored is an element of the conversational score. For example: When we evaluate statements involving physical modality ("can"), we ignore possibilities that violate the laws of nature. And when we evaluate statements involving epistemic modality ("knows"), we ignore possibilities that are known not to obtain. This boundary can be shifted explicitly, as when statements are modified by phrases such as "in view of what is known," or "in view of what custom requires." But this boundary is also governed by a rule of accommodation: if a statement would be false were the boundary to remain stationary, then previously ignored possibilities come into play and make the statement true (Lewis 1979a, 355).

We spent a good deal of the previous chapter looking for principled ways to draw this boundary between possibilities that needed to be ruled out, in order for a knowledge claim to be true, and possibilities that could be ignored. We examined the sensitivity and safety proposals, but found them both less than entirely satisfying. What we see here (and

[59] Although it's tempting to call this latter context a "low-standards" context, I think that would be somewhat misleading, and would prejudice the case against ordinary contexts. Unraised standards might nevertheless be high standards.

what we will see in more detail below) is that we now have something more: we have a proposed conversational mechanism that governs and explains potential shifts in the boundary between possibilities that can be ignored and possibilities that cannot. Lewis illustrates how this works in the case of the skeptical argument in epistemology:

> The commonsensical epistemologist says: "I *know* the cat is in the carton—there he is before my eyes—I just *can't* be wrong about that!" The skeptic replies: "You might be the victim of a deceiving demon." Thereby he brings into consideration possibilities hitherto ignored, else what he says would be false. The boundary shifts outward so that what he says is true. Once the boundary is shifted, the commonsensical epistemologist must concede defeat. And yet he was not in any way wrong when he laid claim to infallible knowledge. What he said was true with respect to the score as it then was. (Lewis 1979a, 355)

Recall the point we noted earlier, namely that standards are more easily raised than they are lowered, and that as a result a shift in which standards are raised according to a rule of accommodation is never fully reversible. Lewis also points out that once the boundary is shifted outward such that additional possibilities must be taken into account, a statement that was perfectly acceptable with respect to the original boundary will never again seem perfectly acceptable—even if the boundary is shifted back to its original position. At best, the statement in question will seem imperfectly acceptable.

This feature of conversational dynamics might give the impression, briefly alluded to above, that contexts in which the boundary has been shifted outward are somehow superior to contexts in which the boundary remains in a more ordinary position. Similarly, we might get the impression that a claim that is true in light of a remote boundary is somehow *truer* than a claim that is false in light of the remote boundary but true in light of a closer boundary. Nevertheless, I follow Lewis (1979a, 355) in seeing no reason to respect this impression. (And in the next chapter I will provide some reasons for *rejecting* this impression.)

Although suggestive, these two moves (i.e., identifying rules of accommodation and resisting the temptation to consider the remote boundary as somehow epistemically superior or prior to closer boundaries) are obviously not enough, at least by themselves, to fully rebut

the skeptical argument. We do have the beginnings of a rebuttal, however—a rebuttal that Lewis fleshes out in his "Elusive Knowledge" (1996).

5.3 Lewis's contextualism: elusive knowledge

Lewis begins with his conviction, which we noted in Chapter 4, that endorsing skepticism and fallibilism are equal parts madness. Forced to choose between these two madnesses, he chooses fallibilism; but it would be better to dodge the choice (1996, 692). Thus, with infallibilism as a starting point—and building on the insights from his (1979a) discussion of relative modality—he proposes and develops the following definition of knowledge:

> Subject *S knows* proposition *p* iff *p* holds in every possibility left uneliminated by *S*'s evidence; equivalently, iff *S*'s evidence eliminates every possibility in which not-*p*. (Lewis 1996, 693)

What does it mean to say that *p* must hold "in every possibility left uneliminated by *S*'s evidence"? What is it for some possibility to be left uneliminated by *S*'s evidence? According to Lewis, "a possibility *w* is *uneliminated* iff the subject's perceptual experience and memory in *w* exactly match his perceptual experience and memory in actuality" (1996, 694). Or, to put it in the active voice: an experience *e* (or memory *m*) with propositional content *p* eliminates *w* iff *w* is a possibility in which *S*'s experience or memory has content different from *p*.[60] Elimination thus consists in a mismatch between *e* (or *m*) and *w*. Strictly speaking, of course, there are very few propositions that hold in *every* possibility left uneliminated by, for example, my evidence. The domain that "every" quantifies over must be restricted in some way—just as the domain must be restricted in some way when I say, "Every student has arrived"—and these restrictions consist in specifying which possibilities can be ignored.

With Lewis's framework now in front of us, the crucial questions are obvious (and familiar): Of all the uneliminated alternative possibilities, which ones may we (and may we

[60] Lewis is careful to avoid claiming that an experience *e,* with propositional content *p,* eliminates *w* iff *w* is a possibility in which *p* is false—because "the propositional content of our experience could, after all, be false" (1996, 694).

not) *properly* ignore? Which of the alternatives are *relevant* to the proposition in question, and as such cannot be properly ignored? To answer these questions, Lewis (1996, 695–98) proposes some rules—three of them prohibitive, four of them permissive—specifying which possibilities we may properly ignore, and which possibilities we may not properly ignore. The prohibitive rules can be briefly summarized as follows.

5.4 Prohibitive rules

The *rule of actuality* tells us that actuality—the subject's actuality—is one possibility that cannot be ignored. The *rule of belief* says that we can ignore neither possibilities that are believed to obtain by the subject, nor possibilities that *should be* believed to obtain by the subject. So far so good. But when we consider the next rule in light of the first two, we run into some complications. According to this third rule, the *rule of resemblance*, we may not ignore any possibility that saliently resembles another possibility that we may not properly ignore (according to the other rules). The problem is that *every* uneliminated possibility ("every" in an unrestricted sense) will resemble actuality in at least one salient respect, namely that of being uneliminated by *S*'s evidence. And of course there are all kinds of far-fetched alternatives (e.g., the Matrix alternative) that, if allowed as relevant, lead inevitably to skepticism. So these possibilities, if we want to avoid skepticism, must be properly ignorable. But given the rule of resemblance, how can making an exception for these possibilities be anything other than *ad hoc*? Lewis admits that he doesn't have an answer to this question; but I'd prefer not to leave it at that. I think we can make some progress on this problem—starting with, perhaps surprisingly, with an objection to Lewis's view.

The objection comes from Stewart Cohen (1998), who argues that Lewis's rule of resemblance cannot after all solve the Gettier problem—because it is a *speaker-sensitive* rule, rather than a *subject-sensitive* rule. We will look at this distinction (between speaker-sensitive rules and subject-sensitive rules) in due course, but first let's examine Lewis's (1996, 696–97) proposed solution to the Gettier problem. Suppose that I believe, for good reasons (e.g., having seen him driving one from time to time), that Nogot owns a Ford, but as it turns out

Nogot does not own the Ford he drives (and in fact doesn't own a car at all). Moreover, I don't have any good reasons to believe that Havit owns a Ford, but as it turns out Havit does in fact own a Ford. In this situation, I have a justified true belief that Nogot or Havit owns a Ford, but I do not have knowledge of this proposition. Lewis's diagnosis of this case is as follows:

> I do not know, because I have not eliminated the possibility that Nogot drives a Ford he does not own whereas Havit neither drives nor owns a car. This possibility may not be properly ignored. Because, first, actuality may not properly be ignored; and, second, this possibility saliently resembles actuality. It resembles actuality perfectly so far as Nogot is concerned; and it resembles actuality well so far as Havit is concerned, since it matches actuality both with respect to Havit's carless habits and with respect to the general correlation between carless habits and carlessness. (Lewis 1996, 696–97)

This, then, is how the rule of resemblance is supposed to solve the Gettier problem: In Gettier cases, there will always be a possibility in which *S*'s justified belief that *p* is nevertheless false, and which saliently resembles actuality. Such a possibility will not be properly ignorable and as such will imply that *S*'s belief is not knowledge.

Let us now examine the distinction between speaker-sensitive rules and subject-sensitive rules. Any given context will include a speaker, a hearer, and a subject (where the same person may fulfill both roles, or even all three roles). We can divide the various rules governing the conversation in that context into two categories.[61] The first category contains rules whose governance depends on facts about the speaker (and perhaps the hearer as well); these rules are speaker-sensitive. The second category contains rules whose governance depends on facts about the subject; these rules are subject-sensitive. (Note again that these categories are not exclusive: a rule can depend on both types of facts, and thus fall into both categories.) Contextualist theories might include both types of rules, but they're notable for emphasizing the role that speaker-sensitive rules play in conversation. (Cohen 1998, 709) But, as Cohen points out, if we pay close attention to this distinction then Lewis's view runs

[61] I use "the conversation" loosely here, which is to say that the conversation could include unspoken judgments or attributions that take place only in the mind.

into troubles. Cohen's (1998, 711) version of a Gettier case is one in which *S* is looking at a sheep-shaped rock on a hill, and as a result comes to believe that there is a sheep on the hill. Unbeknownst to *S,* there is a sheep behind the rock. In this case, *S* has a justified true belief that there is a sheep on the hill, but does not know that there is a sheep on the hill. Lewis would say that *S*'s belief is not knowledge because there is a possibility in which there is a sheep-shaped rock with no sheep behind it, and in this possibility *S*'s belief that there is a sheep on the hill is false. Moreover, this possibility saliently resembles actuality and thus cannot be properly ignored. Again, the two rules that are doing the work here are the rule of actuality and the rule of resemblance. But are these rules speaker-sensitive, or subject-sensitive? The rule of actuality is clearly subject-sensitive, as it is the subject's actuality—*S*'s actuality—that matters. What about the rule of resemblance? Certain possibilities that saliently resemble the subject's actuality must be ruled out; but to whom must these possibilities be salient? As Cohen points out (1998, 710), the possibilities cannot be salient to the subject; otherwise the belief would not be justified and the case would not be a Gettier case. (If the possibility that what appears to be a sheep is merely a sheep-shaped rock [and that the hill is otherwise unoccupied] is salient to *S,* then *S* isn't justified in believing that there's a sheep on the hill.) Thus, the rule of resemblance, unlike the rule of actuality, must be speaker-sensitive.[62]

Cohen argues that this feature of the rule of resemblance provides a way for Lewis to avoid skepticism without resorting to *ad hoc*ery. Recall, as we saw above, that avoiding the skeptical problem, on Lewis's analysis, appears to require an *ad hoc* exception to the rule of resemblance. But Cohen reminds us that it's not mere resemblance between a possibility and actuality that precludes ignoring that possibility; rather, it's *salient* resemblance. And it's part and parcel of any contextualist position that in ordinary contexts, far-fetched skeptical hypotheses are not salient. Thus, the rule of resemblance does not require a troublesome *ad hoc* exception. Lewis made a dialectical concession where no such concession was necessary.

[62] Cohen (1998, 709) uses a lottery case (from Lewis) to show that the rule of resemblance is also subject-sensitive.

So far so good. But Cohen provides a curious diagnosis of Lewis's concession when he puts it in terms of the distinction between subject- and speaker-sensitivity. He claims that an *ad hoc* exception is only required if the rule of resemblance is merely subject-sensitive:

> So in effect, Lewis is forced into his *ad hoc* restriction because he here treats the Rule of Resemblance as if it were merely subject-sensitive. But the rule's speaker-sensitivity enables us to avoid the threat of skepticism without resorting to *ad hoc*ery. (Cohen 1998, 711)

I doubt that a failure to distinguish between subject- and speaker-sensitivity is what leads Lewis to *ad hoc*ery. For it seems to me that if a lack of salient resemblance is what allows a speaker to ignore fantastical scenarios that nevertheless resemble actuality with respect to the subject's evidence, then this lack of salience would also allow the *subject* to ignore such scenarios. It seems to be salience, rather than speaker-sensitivity, that's rescuing Lewis from the *ad hoc* exception, and as a result it's not clear what work the distinction between subject- and speaker-sensitivity is doing here.

Be that as it may, Cohen's examination of the Gettier case does appear to establish the speaker-sensitivity of the rule of resemblance. And this creates the following problem (Cohen 1998, 711). Consider again the Gettier case in which *S* has a justified true belief (but not knowledge) that there's a sheep on the hill. Now introduce *J* into the situation. *J* is looking at the rock from a different angle than *S* is, and as a result can not only see the sheep, but also doesn't see any resemblance between the rock and a sheep. Suppose that *S* expresses to *J* his belief that there is a sheep on the hill. At that point, *J* seems to be in a position to correctly ascribe knowledge to *S*. The possibility that there is a sheep-shaped rock but no sheep on the hill does not saliently resemble actuality, and thus seems properly ignorable by *J*. But, as Cohen points out, it is highly counterintuitive to claim that there is *any* context of ascription in which knowledge is properly ascribed to a Gettierized subject. It seems, in other words, that the speaker-sensitivity of the rule of resemblance renders it ill equipped to solve the Gettier problem.

Are there any other resources that Lewis might bring to bear in light of this limitation on the rule of resemblance? One thought is that the rule of belief might allow us to block the result that *J* can correctly attribute knowledge to *S*. Recall what the rule of belief tells us, namely that we cannot ignore possibilities that are believed or should be believed. This rule is explicitly subject-sensitive (cf. Lewis 1996, 695), but what if we expand its reach to include the speaker as well? Whatever its independent merits, this suggestion will not help Lewis with the Gettier problem. Since *J* doesn't believe that the rock looks like a sheep (from where *S* is standing), the only way for the rule of belief to help is if we can plausibly say that *J* ***should*** believe that the rock looks like a sheep to *S*. And it unfortunately does not seem plausible to say that *J* should believe this. It would be asking too much to require that of *J*.

I do think, however, that consideration of the rule of belief might provide some assistance to Lewis. His view already tells us that *J* cannot ignore what *S* believes or should believe, and that *J* cannot ignore possibilities that saliently resemble actuality. The problem, again, is that the possibility in which there's a sheep-shaped rock on the hill, but no sheep, does not resemble actuality in a way that is salient to *J*. This is what allows *J* to make the mistaken ascription of knowledge. So why not amend the rule of resemblance to say that *J* cannot ignore possibilities that ***should*** saliently resemble actuality? The possibility (call it *q*) in which there's no sheep on the hill, and yet *S* forms the belief that there's a sheep on the hill on the basis of seeing a sheep-shaped rock, is close enough to the actual world that its resemblance to actuality should be salient. (Perhaps a rough way of deciding when a possibility's resemblance to actuality should be salient would be to ask whether the subject's belief is safe—i.e., whether it's true that were the subject to have the belief, it would be true. If not, then nearby possibilities in which his belief would be false will resemble actuality in a way that should be salient.)

It might be objected that this also would be asking too much of *J*. If it would be too demanding to claim that he should believe that the rock looks like a sheep to *S,* then wouldn't it also be too demanding to claim that the resemblance between possibility *q* and

actuality should be salient to *J*? I don't think so, and the reason why involves the element of luck that is present in Gettier-type situations. *S*'s belief that there's a sheep on the hill is only luckily true; it's a matter of luck that, unbeknownst to him, there's an actual sheep behind the sheep-shaped rock. Because his belief is lucky, it doesn't count as knowledge. This doesn't render him epistemically deficient in any way; it's just that a certain amount of epistemic luck precludes knowledge.[63] And we can say something similar about *J*. Given the peculiarity of *S*'s situation (and Gettier cases in general), it is not surprising that "speakers" in slightly different contexts (i.e., contexts in which the evidence that justifies *S*'s belief is not accessible to the speaker, and yet neither is the feature of *S*'s situation that makes his belief lucky) would mistakenly attribute knowledge to *S*. (In a sense, speakers like *J* are the victims of *bad luck*.) Thus it does not seem unreasonable to say that possibilities such as *q* should saliently resemble actuality, and that this is why *J*'s knowledge attribution is incorrect.

Although I think the suggested amendment to the rule of resemblance is reasonable and defensible, it need not ultimately succeed for us to derive benefit from our examination of Lewis's view. For my primary interest in his view—and in particular the conversational mechanism he identifies—is its explanation of how and why the skeptical argument can seem so convincing despite our willingness to deny its conclusion in ordinary contexts. His account of knowledge may need to be revised in various ways, but we can benefit from its explanatory power even without those revisions in hand.

This, then, is where we stand: Lewis would like to avoid making an *ad hoc* exception to the rule of resemblance (in order to solve the skeptical problem). Cohen suggests that the salience criterion, which is already built into the rule, makes the *ad hoc* exception unnecessary. Unfortunately, this same criterion renders the rule powerless against the Gettier examples. For the rule is speaker-sensitive, and if we allow possibilities featuring non-salient resemblances to be ignorable then there will be some Gettier cases in which knowledge can be properly attributed to the subject—surely the wrong result. I agree with

[63] See Pritchard (2005) for an extended treatment of epistemic luck. He describes the kind of luck that characterizes Gettier cases as "veritic luck."

Cohen that the rule of resemblance must depend on facts about the speaker, and that this speaker-sensitivity renders the rule ill-equipped to handle the Gettier problem. Thus I proposed that we revise the rule of resemblance to say that possibilities that *should* saliently resemble actuality are not properly ignorable. Speakers such as *J* are victims of bad luck insofar as there are some resemblances that should be salient to them but are not.

This completes our discussion (and supplementation) of Lewis's prohibitive rules. We now move on to the permissive rules—i.e., the rules that tell us which possibilities *can* be properly ignored.

5.5 Permissive rules

The first permissive rule is the *rule of reliability* (Lewis 1996, 697). This rule tells us that we can properly ignore a possibility in which a generally reliable process (e.g., perception, memory, or testimony) isn't working properly. Thus in the case of perception, the rule of reliability produces something like the following principle:

> (1) If *S*'s perceptual system represents an object *x* as *F* (where *F* is a perceptible property), and this causes or sustains in the normal way *S*'s belief of *x* that it is *F,* then we can properly ignore a possibility in which *x* is not *F* (cf. P. J. Graham 2006, 95).[64]

As Lewis points out,

> it is possible to hallucinate—even to hallucinate in such a way that all my perceptual experience and memory would be just as they actually are. That possibility never can be eliminated. But it can be ignored. And if it is properly ignored—as it mostly is—then vision gives me knowledge." (1996, 697)

[64] Graham formulates this principle (and related principles) in terms of justification rather than ignorable possibilities, but since Lewis eschews talk of justification (and in fact claims that on his account justification is neither necessary nor sufficient for knowledge [although Cohen {1998, 717n6} disputes this]) I have taken the liberty of reformulating the principle.

What (1) tells us, then, is that as we partake in our day-to-day perceptual commerce, we can typically ignore possibilities in which, for example, we're hallucinating. And the same goes, *mutatis mutandis,* for memory, testimony, and other generally reliable processes.

Lewis is quick to point out, however, that the propriety of such ignorings is defeasible. The rule of reliability can be defeated by the rule of actuality. If *S* is actually hallucinating, then the epistemic principle in (1) is overridden by the rule of actuality. And similarly for testimony: if *S* is the recipient of false testimony, then the possibility that what he's hearing is false also happens to be actuality, and as such is not a possibility that can be properly ignored. The rule of reliability can also be defeated by different combinations of the prohibitive rules. For example, as we saw above (albeit with some qualifications), the rule of reliability is defeated in Gettier cases by the familiar combination of the rule of actuality and the rule of resemblance. When *S* is in fake barn country, the rule of reliability is defeated by the fact that there are too many barn façades in the neighborhood. His perceptual faculties are working just fine, but there are too many possibilities that closely resemble actuality in which his belief (that he's looking at a barn) is false. The next two permissive rules are both rules of method. The first says that we can presuppose that a sample is representative, and the second says that we can presuppose that the best explanation of our evidence is also the right explanation (Lewis 1996, 697). In other words, we can properly ignore possibilities in which a sample is unrepresentative, and we can properly ignore possibilities in which the right explanation is something other than the best explanation—assuming, that is, that none of the prohibitive rules are in effect. And finally, there is the *rule of conservatism*: if a possibility is commonly ignored (and moreover it's *common knowledge* that this possibility is commonly ignored), then it can be *properly* ignored. (This rule, like the other permissive rules, is defeasible.) Lewis also points out a "triviality" that could be considered an additional rule; call it the *rule of attention* (1996, 698). This rule simply states that if a possibility is attended to, then it isn't properly ignorable.

I would like to highlight one final feature of Lewis's modal analysis of knowledge before we examine some objections and crystallize some insights for our own purposes.

Notice that his analysis allows us to maintain our commitment to epistemic closure. And not only that—it also provides a diagnosis of what's going on in, for example, the lottery case (Lewis 1996, 702). We saw that the lottery case (not to mention cases involving situations much more mundane than winning the lottery) provided a temptation to deny closure. Lewis's insight with respect to such cases is that there is a shift in context as we move from one aspect of the case to the next. When I am considering the question of whether I'll be able to afford to go on an African safari next summer, I can (and do) know that I won't. The possibility that I'll win the lottery between now and then is not salient, and as such can be properly ignored. But once the possibility that I'll win is on the table, it is no longer being ignored and thus can no longer be *properly* ignored. By mentioning the chance that I'll win, remote as it may be, I have effected a shift in context and brought a previously properly ignorable possibility back into play. In this new context, a claim to know that going on safari won't be affordable next year would be false. Thus there is no single context in which I'm both forced to claim knowledge that I won't be able to afford to go on safari *and* forced to deny knowledge that I won't win the lottery. (And the same goes for other ordinary propositions with lottery-style entailments.)

5.6 Responding to the skeptic

Now we are in a position to ask: How does the conversational mechanism that Lewis identifies in his (1979a) and develops in his (1996) provide us with an answer to the skeptic? Thanks to Cohen (1998), we've already seen, at least in broad outlines, how this is supposed to work. But I'd like to examine it in a bit more detail, and with specific reference to the skeptical argument as formulated in Chapter 1 and discussed at length in Chapter 4. Here, once again, is the argument that generates the skeptical challenge:

> (2) If I know that I had pancakes for breakfast, and that my having pancakes for breakfast entails my not being plugged into the Matrix, then I know that I'm not plugged into the Matrix. (Epistemic closure principle)

(3) I know that my having pancakes for breakfast this morning entails that I'm not plugged into the Matrix. (Premise)

So, (4) If I know that I had pancakes for breakfast this morning, then I know that I'm not plugged into the Matrix. (2, 3)

(5) But I don't know that I'm not plugged into the Matrix. (Premise)

So, (6) I don't know that I had pancakes for breakfast this morning. (4, 5)

If Lewis's analysis (plus Cohen's addendum) is right, then we should be looking for an explanation of the force of the skeptical argument in terms of a possibility that is typically properly ignored but has become salient in the current context. And this is exactly what we find. In the vast majority of contexts in which I claim to know that I had pancakes for breakfast, this claim is true. But there are various possibilities—e.g., the possibility that I'm plugged into the Matrix—that resemble actuality with respect to my evidence for what I had for breakfast this morning. In those contexts in which I do know what I had for breakfast, that is (in part) because I am properly ignoring Matrix-type possibilities. But as soon as such a possibility is brought to my attention, which is to say as soon as I start attending to it, the rule of attention tells us that I can no longer ignore it. And this insight is reflected in the epistemic closure principle—which constitutes the first premise of the skeptical argument. So the contextualist response to this argument is to acknowledge, first of all, that once we've started down the skeptical path (by entertaining the first premise), we are indeed inexorably led to the skeptical conclusion in (6). But we need not travel that path at all—it is, after all, an unusual and for most purposes not particularly useful path. In ordinary, everyday contexts (which might be in many respects quite epistemically sophisticated), we simply do not need to entertain the skeptical doubts that, once entertained, do indeed undermine our knowledge.

Recall the two crucial points that we first encountered in Lewis's (1979a). First: A rise in standards for the correct attribution of a concept does not imply that previous attributions, in different contexts with lower standards, were in any way incorrect or

unacceptable. And second: Higher standards in a particular context do not imply any sort of superiority for that context. In other words, the attribution in the lower-standards context, if correct, is no less robust or appropriate than the attribution in the higher standards context.

These affirmations lead us to a third point, which prefigures one of the main lessons of the next chapter. One of the virtues of epistemological contextualism is that it preserves ordinary knowledge attributions while providing a plausible explanation of why the skeptical argument is so compelling. Lewis's account is "built to explain how the skeptic manages to sway us—why his argument seems irresistible, however temporarily" (1996, 699). This feature of his account suggests a desideratum for responses to the skeptical argument—and for responses to structurally parallel arguments, such as the consequence argument. The desideratum is that any response that rejects the conclusion should also provide an explanation of the force of the argument. A proposed response to the argument will be commendable only if, and insofar as, it provides a plausible explanation of why the argument is so compelling.

The development of just such a response is what I undertake to do in the next chapter. But first, we must consider some objections to the contextualist response to skepticism.

5.7 Objections to epistemological contextualism

The contextualist maintains that many of our ordinary knowledge attributions are true. At the same time, the contextualist explains the force of the skeptical argument by positing a conversational mechanism that raises the standards for knowledge claims in certain contexts. I think the spirit of this move can be adopted and incorporated into a response to the consequence argument; but before I do so, we need to look at some objections to the move.

I will focus on four objections—which claim, respectively, (a) that contextualism is unsatisfying because it is a theory of knowledge *attributions* rather than a theory of knowledge; (b) that contextualism forces us to claim, implausibly, that whether or not some

subject knows depends on the attributor's context; (c) that contextualism wreaks havoc with norms of assertion and practical reasoning; and (d) that contextualism is committed to a problematic form of *semantic blindness.*[65] These are of course not the only objections in the literature, but I will treat them as representative: the first two because they are to my knowledge the most common; and the second two because they are put forward by a friend of contextualism, namely Hawthorne. Although Hawthorne is not himself a contextualist about 'knows' (see his 2004), he does advance a contextualist account of freedom (2001), which we will consider in the next chapter. In addition, the semantic blindness objection is important because an appeal to semantic blindness is considered by many opponents of contextualism to be the best response to a host of other objections (and yet wildly implausible).[66] Thus I will consider the semantic blindness issue as something of a consolidated objection. If I can successfully respond to these representative objections, then it will be safe to treat the view as at least provisionally defensible.[67]

[65] In Chapter 6 we will also consider three objections from Feldman (2004). And I should note that Cappelen and Lepore (2005) have mounted a vigorous critique of contextualism in general (i.e., not just in epistemology). On account of limited space, and since their arguments apply more directly to the sort of contextualism advanced by DeRose (1995, 1999, 2002, 2005, 2009), I will set them aside for now and instead refer the reader to Hawthorne (2006).

[66] DeRose (2009, 154) lists the objections that raise the problem of (i.e., invite a response that appeals to) semantic blindness. First, the objection from comparative judgments of content: the denial of knowledge in a high-standards case appears to contradict the attribution of knowledge in a low standards case. Second, the objection from metalinguistic claims: two speakers, in two different contexts (again, one of which is a high-standards context and one of which is a low-standards context), would likely have to claim that each other's claims are false. Third, the objection from belief reports: that the speaker in a high-standards context, while himself denying knowledge, would claim that the speaker in the low standards contexts believes that knowledge should be attributed (and vice versa).

[67] And it's worth reiterating here that the conversational mechanism that forms the core of epistemological contextualism will ultimately be deployed in an explanatory role; thus the first two objections do not apply directly to my own incorporation of the contextualist maneuver. Nevertheless, rebutting these objections will lend additional plausibility to my project.

Objection: not a theory of knowledge

Perhaps the most general and fundamental objection to contextualism is that it is not a theory of knowledge (as it purports to be), but rather a theory of knowledge attributions. The claim is that contextualism in a sense misses the point because contextualists are talking about something different than what epistemologists are talking about. But this charge need not stick, for contextualist theories in epistemology are not *merely* theories of knowledge attributions. It is true that these theories are characterized by a claim about knowledge attributions: the claim that, for some speaker *A* who says of some subject *S,* "*S* knows that *p,*" how strong an epistemic position *S* must be in with respect to *p* for *A*'s assertion to be true can vary according to features of *A*'s conversational context (cf. DeRose 1995, 670). But the contextualist still can (and should) say that *A*'s assertion that '*S* knows that *p*' is true iff *S* knows that *p*. Thus the contextualist can claim that the standards for knowledge (or, more precisely, for what "counts as knowing," as we will see below) vary according to features of *A*'s conversational context, and thus that his theory is a theory of knowledge and not merely a theory of knowledge attributions.

Unfortunately, it appears that in answering the first objection we have now committed the contextualist to what strikes many as implausible: the claim that whether *S* knows that *p* depends on *A*'s context. So we will have to sharpen up our language in order to respond to this second objection.

Objection: implausible commitments

This objection is also prompted by some apparently incautious remarks from Lewis. First, he considers the possibility that epistemological theorizing precludes its own ends:

> Maybe epistemology is the culprit. Maybe this extraordinary pastime robs us of our knowledge. Maybe we do know a lot in daily life; but maybe when we look hard at our knowledge, it goes away. ... Maybe ascriptions of knowledge are subtly context-dependent, and maybe epistemology is a context that makes them go false. Then epistemology would be an investigation that destroys its own subject matter. (1996, 692)

Later he explains how this works when epistemology is conceived of as he conceives of it, namely as an investigation of ignoring possibilities:

> Unless this investigation of ours was an altogether atypical sample of epistemology, it will be inevitable that epistemology must destroy knowledge. That is how knowledge is elusive. Examine it, and straightway it vanishes. (1996, 698)

If this is what happens when epistemologists investigate their own knowledge, then presumably this is also what happens when epistemologists investigate the knowledge of others. And *that* might lead someone to object that the contextualist is committed to claiming that whether or not the subject knows depends on the attributor's context. But, as DeRose (2009, 212–17) points out, the contextualist need not be committed to this (admittedly implausible) claim. Instead, what the contextualist is saying about shifts in context is the following (cf. DeRose 2009, 215–16):

> (7) First, *S* was such that the proposition expressed about her by the sentence "*S* knows that *p*" in *A*'s conversation was true of her. But then, because *A*'s context changed so that "*S* knows that *p*" came to express a more demanding proposition, *S* was such that the (new) proposition that would have been expressed by "*S* knows that *p*" in *A*'s context was not true of her.

It is thus not accurate, *pace* Lewis, to say that whether *S* knows depends on *A*'s context—that *S* has somehow lost her knowledge when *A*'s context changed. Rather, what *S* has lost is a certain *relation to A*'s *context* (cf. DeRose 2009, 214). *S* once met the standards governing *A*'s conversation, but now does not. Now, admittedly, (7) is cumbersome and tedious. Thus DeRose recommends the following as a convenient shorthand:

> (8) First *S* met the epistemic standards set by *A*'s context and then, because *A*'s standards went up, she failed to meet the standards set by *A*'s context

and the following as super-shorthand:

(9) First *S* counted as knowing, and then she didn't.

And (9) is different, albeit subtly so, from saying that first *S* knew, and then she didn't—that whether *S* knows depends on *A's* context. Thus if the contextualist is careful to use the right shorthand, he can avoid the implausible claim that has been attributed to him.

Notice also that this still seems to be a theory of knowledge: it's a theory of when a certain epistemic position is strong enough to bear a certain relation to certain standards for knowledge. It is perhaps an unusual theory of knowledge, and it requires a bit of work to keep track of things, but it remains a theory of knowledge nonetheless.

Objection: assertion and practical reasoning

Does contextualism lead to unacceptable results when it comes to assertion and practical reasoning? This is Hawthorne's complaint against the view, which can be summarized as follows (cf. Hawthorne 2004, 85–91; my presentation also borrows from Feldman 2007). We have seen that certain features of the speaker's context determine whether or not knowledge can be properly ascribed to a subject (who of course in some cases is identical with the speaker). But when we consider assertion and practical reason, it seems clear that the appropriateness of these two activities cannot depend on facts about the speaker's context (if the speaker is in a different context than the subject); instead, it must be features of the *subject's* context that matter to assertion and practical reasoning. Consider two examples (which I will expand upon below). In an ordinary context it seems perfectly appropriate for some *S,* who had pancakes for breakfast this morning, to assert that she had pancakes for breakfast. The fact that I am now entertaining the Matrix hypothesis does not render her assertion inappropriate. Similarly (as *S*'s thoughts turn to lunch), given that *S* parked her car in Lot F (which is a long walk from her office), it seems perfectly appropriate for her to use as a premise in her practical reasoning about lunch the fact that her car is parked in Lot F.

My entertaining of the possibility that her car has been stolen since she parked it there this morning does not render her practical reasoning somehow invalid.

So far so good. But recall the knowledge norm of assertion, which we briefly considered in the previous chapter. According to the knowledge norm of assertion, which enjoys at least *prima facie* plausibility, we should only assert what we know. And there is a related normative constraint on practical reason—which appears to enjoy roughly the same degree of *prima facie* plausibility, and which tells us that we should only use propositions that we know as premises in our practical reasoning. Now note that if the contextualist adopts these constraints on assertion and practical reasoning, then he is forced to countenance situations in which knowledge and assertion (or knowledge and practical reason) come apart. Imagine that a subject *S* is in an ordinary (low-standards) situation and asserts that she had pancakes for breakfast (which happens to be true). We can easily suppose that *S* knows that she had pancakes for breakfast (or, to put it more carefully, that someone in *S*'s context who attributes knowledge of that proposition to *S* would do so correctly), and thus that her assertion adheres to the knowledge norm of assertion. But now an epistemologist *E* is considering the question of whether *S* knows that she had pancakes for breakfast. The Matrix possibility is very much in play, let us suppose. *E* concludes that *S* doesn't know that she had pancakes for breakfast. He admits, though, that her assertion to that effect was nevertheless appropriate. What *E* is committed to, then, is something like the following conjunction:

(10) *S* doesn't know that she had pancakes for breakfast, but she was licensed to assert that she had pancakes for breakfast.

This appears to be a blatant violation of the knowledge norm of assertion.

A similar problem arises with respect to practical reasoning. Suppose that *S* is deliberating about whether to go out for lunch, or whether to eat on campus. She reasons as follows: "Going out for lunch would be nice, but my car is parked in Lot F—which is a bit of

a hike—and I have a few things to take care of before my 2:00 pm class; so I'll just eat on campus today." In this scenario, *S* has used "My car is parked in Lot F" as a premise in her practical reasoning. And this seems fine, because it is not hard to imagine that she knows that her car is parked in Lot F. (As before, it would be more accurate to say that someone in her context could correctly attribute knowledge of that proposition to her.) But when our epistemologist *E* considers her situation, he cannot help but consider the possibility that her car has been stolen since she parked it there. He concludes that she doesn't know that her car is parked in Lot F. What *E* is committed to, then, is something like the following conjunction:

(11) *S* doesn't know that her car is parked in Lot F, but it was appropriate for her to use "My car is parked in Lot F" as a premise in her practical reasoning.

This appears to be a blatant violation of the knowledge norm of practical reasoning.

How might the contextualist respond to these criticisms? The obvious response, of course, would be to deny the respective knowledge norms. I am sympathetic to this response (particularly as it pertains to the knowledge norm of assertion, as alluded to in the previous chapter), but I think our response will be stronger if we can maintain neutrality with respect to the knowledge norms. Richard Feldman—no contextualist himself—provides just such a response (2007, 216–18). He points out, in the spirit of the maneuver we considered in the previous subsection, that the knowledge norms can be reformulated metalinguistically, and provides a formulation of the assertion principle that draws inspiration from Cohen (2004):

(12) *S* is not to be criticized for asserting *p* just in case "*S* knows that *p*" is true in *S*'s context. (Feldman 2007, 218)

Feldman's defense of principle (12) seems largely right to me, and provides a helpful segue into the next objection. Here's what he has to say:

> If, as contextualists admit, the context sensitivity of K-sentences [i.e., sentences of the form, "S knows that *p*"] is not something that we regularly recognize, then it is likely that we will not notice some consequences of that context sensitivity. Of course, if the data on which contextualism is founded are to be believed, our reactions to K-sentences themselves do change with context. But our responses may not be attuned to the metalinguistic assertion principle because we do not realize that we are making these shifts. If it is reasonable to believe that we are to this extent *unaware of what we are doing,* it is difficult to see what is additionally troublesome by our failure in this regard. Thus, the fact that an acceptable assertion principle will be metalinguistic does not seem to me to be a new and independently implausible commitment of the theory. It is just a further implication of a more familiar point. (Feldman 2007, 218, emphasis mine)

The phenomenon that Feldman is referring to here—the fact that we may not be aware of contextual shifts, and as a result may not notice some of the consequences of context-sensitivity)—is commonly referred to as "semantic blindness." Some have thought that semantic blindness poses the most serious problem for contextualism (cf. Ichikawa 2009), and what we have learned here is that Hawthorne's objection to contextualism is only as powerful, and may even collapse into, the objection from semantic blindness. To that objection we now turn.

Objection: semantic blindness

Cohen provides a nice summary of what Hawthorne (2004, 107–11) has dubbed "semantic blindness":

> A central feature of the contextualism I defend is that we mistakenly assume certain knowledge ascriptions conflict which in fact do not. … I am committed to the view that, although ascriptions of knowledge are context-sensitive, competent speakers can be unaware of this, and so can be misled by it. Although their knowledge ascriptions track the shifts in context, they are unaware that these shifts are occurring. (Cohen 2010, 121–22)

Or, to put the point in more Lewisian terms, the idea here is that speakers can be competent with respect to the conversational rules that govern the raising and lowering of the standards for knowledge, but without being aware of precisely what it is that their competence consists in. Is this a problem for contextualism? Some (e.g., Hawthorne 2004) have suggested that it is; that the semantic blindness thesis is implausible enough to present problems for any view

that's committed to it. Cohen's response (2010, 122) is to argue that there are other, less controversial examples of semantic blindness, which should alleviate concerns about its manifestation in a contextualist account of knowledge. His example is the familiar one, from Unger (1979), of *flatness.* Ascriptions of flatness do indeed seem to be context-sensitive, and yet competent speakers can easily be misled into thinking that ascriptions of flatness made in low-standards contexts conflict with flatness ascriptions made in higher-standards contexts. If this is right, then there is at least a precedent for semantic blindness. Cohen admits, however, that there is a disanalogy between flatness and knowledge—namely that there is very little resistance to the claim that flatness ascriptions are context-sensitive, whereas there is stiff resistance to the claim that knowledge ascriptions are context-sensitive. We have, it seems, reached another stalemate.[68]

But on this particular point, I think we can do better than a stalemate. There are two considerations that tip the balance in favor of the contextualist. The first is a feature of Lewis's analysis of knowledge that we haven't yet made explicit. He claims that "the link between knowledge and justification must be broken," and more specifically that justification is not necessary for knowledge:

> What (non-circular) argument supports our reliance on perception, on memory, and on testimony? And yet we do gain knowledge by these means. And sometimes, far from having supporting arguments, we don't even know how we know. We once had evidence, drew conclusions, and thereby gained knowledge; now we have forgotten our reasons, yet we still retain our knowledge. Or we know the name that goes with the face, or the sex of the chicken, by relying on subtle visual cues, without knowing what those cues may be. (Lewis 1996, 692–93)[69]

Although the phenomenon here is perhaps more general than mere semantic blindness, this is another example of language users being competent without having a grasp of precisely what it is that constitutes that competence. So it seems that we have an uncontroversial

[68] See Chapter 2, §2 for a brief discussion of dialectical stalemates.

[69] Although, as noted above, Cohen (1998, 717–18n6) disputes Lewis's claim on the grounds that Lewis's notion of justification is overly restrictive.

example of semantic blindness (viz., flatness ascriptions) and some (more or less) uncontroversial examples of a different sort of blindness that pertains to self-ascriptions of knowledge. This makes it less surprising that knowledge would admit of semantic blindness as well.

The final consideration, which further weakens the semantic blindness objection, comes from Keith DeRose (2010, 159–60). He points out that invariantists (i.e., those who would deny that the standards for knowledge-ascription vary according to context) must face their own version of the semantic blindness problem. Start with flatness. It's plausible to claim that flatness ascriptions are context-sensitive, but suppose for a moment that they are actually invariant—which means that there is only one standard for flatness. Even on this invariantist picture, it still seems true to say that my glass table is flatter than my driveway. And I see no reason to reject the assumption that nothing can be flatter than something that's already flat. Thus, on the invariantist picture, my driveway is not after all flat, which is to say that attributing flatness to it would be inappropriate. But if we ask competent language users whether my driveway is flat, they will probably answer in the affirmative. In other words, the invariantist about flatness will be forced to say that, although ascriptions of flatness are not context-sensitive, competent language users can be unaware of this, and even misled by it. So the flatness invariantist faces a semantic blindness problem.

Does the invariantist about knowledge face the same problem? It appears that he does. If his standards for knowledge are infallibilist, and require that I be able to rule out the possibility that I'm plugged into the Matrix (in order to know that I had pancakes for breakfast this morning), then clearly there's a semantic blindness problem. For competent language users will no doubt affirm—incorrectly, according to the infallibilist account of knowledge—that I know that I had pancakes for breakfast. But even if the standards for knowledge are not so strict, the problem remains. For example, on a sensitivity theory of knowledge, my belief that I'm not plugged into the Matrix is not sensitive, and hence not knowledge. But, again, most competent language users would claim that I know that I'm not plugged into the Matrix. So the sensitivity theorist is forced to posit semantic blindness as

well. What about the safety theorist? Recall the red barn case from the previous chapter, in which Smith sees a red barn in a landscape full of blue barns. If Sosa's diagnosis of the case is correct, then Smith does not know that he's looking at a red barn, because that belief is inferred from at least one other belief that isn't safe (i.e., the belief that he's looking at a barn). And yet there will be plenty of competent language users who would claim that Smith knows that he's looking at a red barn. Even the safety theorist has to deal with the occasional bout of semantic blindness.

I conclude that the contextualist move in epistemology survives the representative challenges that we have entertained. With contextualism thus defended, we are now in a position to put it to use. More specifically, we are in a position to take the meta-semantic move that the contextualist makes and incorporate it into a novel response to the consequence argument.

• Chapter 6 •

Fundamentalist contextualist compatibilism

6.1 Introduction

In the previous chapter we briefly considered Lewis's claim that there is no reason to respect the impression that high-standards contexts are somehow superior to low-standards contexts—no reason to respect the impression that a claim that is true in light of a remote boundary is somehow *more* true than a claim that is false in light of the remote boundary but true in light of a closer boundary. In this chapter, I would like to go a little bit further by providing positive reasons to reject the primacy of high-standards contexts. These reasons will lay the groundwork for, and serve as the foundation of, a novel response to the consequence argument.

6.2 Southern fundamentalism

George Graham and Terence Horgan (Horgan and Graham 1991, Graham and Horgan 1994) have recently defended a version of realism about folk psychology that they call *southern fundamentalism*.[70] Their fundamentalist version of realism differs from other versions in that they reject a certain epistemic principle that is endorsed by traditional realists and anti-realists alike. Before we discuss this epistemic principle, however, we should briefly examine

[70] Not to be confused with *liberal fundamentalism* (cf. P. J. Graham 2006, 2007), which is a version of the view that testimony-based beliefs are epistemically direct.

what's at stake in this debate and how it relates to the skeptical and consequence arguments. The question of whether to be a realist or anti-realist about folk psychology is a question about whether humans are "true believers"—about whether "humans generally do undergo the [folk psychological] events, beliefs, desires, and so forth that we normally attribute to them" (Horgan and Graham 1991, 107). Realists claim that we are true believers, whereas anti-realists (or eliminativists) deny that we are true believers. In other words, anti-realists claim that folk psychology is radically and categorically false. The dispute between these two factions typically revolves around whether humans meet certain conditions that are taken to be required for humans to be true believers. (Horgan and Graham call these "putative true-believer" conditions, or PTBs.) For example, it is typically claimed that in order for humans to be true believers, folk psychology must be absorbable into mature science. Moreover, anti-realists will claim (and most realists will agree) that evidence against the scientific absorbability of folk psychology is evidence against realism, i.e., evidence against humans being true believers.

Graham and Horgan turn this epistemic principle on its head, claiming that evidence against, e.g., the scientific absorbability of folk psychology is not evidence against true believerhood, but rather evidence against absorbability being a condition on true believerhood. This epistemic principle of evidential dynamics is the linchpin of southern fundamentalism, and is in fact the third of the view's three main tenets—which are summarized as follows (Horgan and Graham 1991, 109):

(1) Humans are true believers.

(2) The thesis that humans are true believers is enormously well warranted, on the basis of total current evidence.

(3) For each PTB condition C, if there were to arise strong epistemic warrant for the thesis that humans do not satisfy C, then (i) this would thereby confer strong epistemic warrant upon the thesis that C is not really a prerequisite for being a true believer, and (ii) it would not confer any significant degree of epistemic warrant upon the thesis that humans are not true believers.

The view that I'll be developing in this chapter is, as it were, a version of southern fundamentalism about free will. Instead of arguing that humans are true believers, I'll be arguing that humans are *free agents*—i.e., that some of the actions humans perform are free actions (for which they're morally responsible). Toward this end, I will construe the dispute between compatibilists and incompatibilists as revolving around the question of whether we satisfy certain conditions that are taken to be required for us to be free agents. (I will call these "putative free agent" conditions, or PFAs.) In particular, I will be concerned with one condition, which becomes relevant when we consider the deterministic hypothesis: the condition that we be able to do otherwise than we actually do, holding fixed the actual past history of the world and the actual laws of nature. Thus free will fundamentalism, like Horgan and Graham's southern fundamentalism, can also be characterized by three central tenets:

(4) Humans are free agents.

(5) The thesis that humans are free agents is enormously well warranted, on the basis of total current evidence.

(6) For each PFA condition C, if there were to arise strong epistemic warrant for the thesis that humans do not satisfy C, then (i) this would thereby confer strong epistemic warrant on the thesis that C is not really a prerequisite for free agency, and (ii) it would not confer any significant degree of epistemic warrant on the thesis that humans are not free agents.

This position is perhaps more controversial than southern fundamentalism—for example, it seems that the third tenet (6) requires more defense than its southern fundamentalist counterpart—but it is nonetheless the right position to hold. Or so I shall argue.

6.3 Austerity vs. opulence

Before I can argue for my own contextualist version of free will fundamentalism, I need to introduce a couple of additional terms. First, Horgan and Graham point out that there are two conceptual tendencies in this debate: one toward what they call *austerity*, and one toward what they call *opulence*. These tendencies lead to, respectively, either an austere conception of folk psychology or an opulent conception of folk psychology. And the difference between these two conceptions has to do with the alleged metaphysical gap between *resonant intentional systems* and true believers.

Resonant intentional systems, according to Horgan and Graham (using some terminology adapted from Dennett) are systems "whose behavior can be usefully predicted and [apparently truly] explained by ascribing to them beliefs, desires, and related attitudes" (1991, 113). What makes these intentional systems *resonant* is that their "overall behavioral repertoire is sufficiently rich, environmentally intricate, and *prima facie* rational that under *ordinary, behavior-based*, epistemic standards for attribution of folk psychological attitudes," nobody would question whether such systems have the relevant attitudes (Horgan and Graham 1991, 113). Moreover, the paradigm RIS is also a competent language user. To borrow their examples (1991, 113), it seems that Quine is clearly an RIS, whereas "the moth on his copy of *The World as Will and Idea*" is clearly not. And finally, recognizing that there are various fanciful scenarios in which robots or other non-RISs, if under the control of sophisticated individuals or systems, could *appear* to be RISs, Horgan and Graham distinguish between puppet RISs and non-puppet RISs.

All parties to the debate agree that humans are RISs. But anti-realists deny that being an RIS is sufficient for being a true believer; they think that there's a wide gap between RIS-hood and true believerhood. (This gap, we might say, is partially constituted by the scientific absorbability of folk psychology and the language of thought.) Fundamentalists, on the contrary, think the gap is "metaphysically negligible" (Horgan and Graham 1991, 115); all you have to be, in order to be a true believer, is a non-puppet RIS. Thus, the difference between the fundamentalist (austere) conception of folk psychology and the opulent

conception of folk psychology has to do with how much is required, on top of being non-puppet RISs, in order for humans to be true believers.[71] Fundamentalists austerely claim that very little is required, whereas secularists opulently claim that significant conditions need to be satisfied. Or, to put the distinction slightly differently, the austere concept of a true believer includes fewer conditions, or prerequisites (basically only one—being a non-puppet RIS—depending on how you individuate conditions) than the opulent concept of a true believer.

Horgan and Graham (1991, 116) provide a helpful example of an austere concept: that of *being able to fly*. Certain creatures are clearly able to fly, based on our ordinary, behavior-based, epistemic standards for attributions of flying ability. Now some sophisticated scientific theorist might come up with a condition, C, that, she purports, is required for a creature to be able to fly (e.g., C could be a minimal ratio of surface area to body weight). If we then discover a creature that can fly (according to our ordinary standards for such attributions), but fails to meet the putative condition C, what is our reaction? We don't give up on our attribution of flying ability; rather, we acknowledge that C must not be required for being able to fly after all. Horgan and Graham are claiming that the concept of *being a true believer* is in this respect analogous to the concept of being able to fly.

Before we examine the next step in the development of Graham and Horgan's view, I would like to pause to consider how this distinction between austerity and opulence applies to the concept of *acting freely*.

[71] I am construing the debate here as between fundamentalists and anti-realists, but Horgan and Graham (1991) actually develop their "southern fundamentalism" as an alternative to both the "Western secularists" (i.e., the anti-realists) and the "Eastern churchmen" (i.e., the realists who affirm the putative true believer conditions as requirements on true believerhood). For my purposes the important question is whether we should affirm the PTBs, so I am ignoring the distinction between Western secularists and Eastern churchmen.

6.4 Conceptions of acting freely: austere vs. opulent

As we apply this distinction to the concept of acting freely, I think it will be helpful to remind ourselves of what we are asking when we ask whether, for example, my raising of my coffee cup was a free action. Assuming, as we are, that acting freely requires the ability to do otherwise, we are asking whether, given that I have raised my cup, it was true that I could have refrained from raising my cup. In other words, just prior to the cup-raising, was the following can-claim true or false?

(7) I can refrain from raising the cup.

Thus, another way of asking whether (and what it means to say that) the concept of acting freely is austere or opulent is asking whether such can-claims are true or false.

There are certain contexts in which it is clearly true that I can refrain from raising my cup. For example, in *this* context, in which I am sitting at my desk and typing on the computer, and pausing occasionally to have a sip of coffee, (7) seems obviously true—at least according to our ordinary, behavior-based, epistemic standards for attribution of abilities such as that of being able to refrain from raising a cup. In fact, it's difficult to imagine a scenario in which (7) would be false; perhaps it would be false if I had somehow fastened the cup to my hand, such that every time I moved my hand I was thereby moving the cup. At any rate, the salient point here is that in asking whether (7) is true, there is typically no thought of the past history of the world, or of the laws of nature—much less of whether I can refrain, given that I don't refrain, and holding fixed the actual past and the laws of nature prior to my raising of the cup. This condition on the truth of (7)—which we might summarize as the requirement that I be able to refrain from raising the cup, where that refraining is an extension of the actual past—is thus going to be part of the opulent conception of acting freely. (There may be other opulent conditions as well, but this is the relevant one for our purposes.)

Another way, and perhaps a clearer way, to apply this distinction to the concept of acting freely, is to construe it in Perry's (2004) terms, as discussed in Chapter 2. Recall, first, that a proposition is *settled* if and only if it is entailed by other propositions that have been made true (either by events, or by something else). Recall also that Perry distinguishes between a weak account of ability and a strong account of ability. The question that divides these two accounts of ability is the following: Can some agent *S* perform or refrain from some action *A,* despite the fact that whether she will perform or refrain has been settled ahead of time? (Perry 2004, 237) An affirmative answer represents a weak view of ability, whereas a negative answer represents a strong view of ability. Given the strong view's negative answer, we can say that a theory of action, which takes a strong view of ability, incorporates the following principle (i.e., the following condition on acting freely):

(8) If S can perform A at *t,* then at no time earlier than *t* is it settled whether S performs A at *t*.

According to the austere conception of acting freely, (7) is true at the time of my cup-raising, even if the world is deterministic. This implies that I can refrain from raising my cup even though it is settled beforehand that I raise it. In other words, the austere concept of free action implies a weak account of ability (in Perry's sense). The converse isn't true, for there could be additional conditions on acting freely even if one were to adopt a weak account of ability. Nevertheless, for the purposes of responding to the consequence argument, the relevant condition on acting freely is the requirement that agents be able to do otherwise as an extension of the actual past. Thus, I will treat the austere concept of acting freely and the weak account of ability as roughly synonymous. The ascriptions, or attributions, that Horgan and Graham are primarily concerned with are ascriptions of folk psychological attitudes: beliefs, desires, intentions, and so on. The ascriptions that we'll be primarily concerned with are ascriptions of freedom. But since it's more common, and more natural, to ascribe abilities (i.e., to formulate and evaluate can-claims) than it is to ascribe freedom of action, I

will also treat such ascriptions as basically interchangeable. That is to say, an ability attribution that rests on a weak account of ability will be taken to represent an austere conception of freedom, and an ability attribution that rests on a strong account of ability will be taken to represent an opulent conception of freedom.

With the distinction between austerity and opulence firmly in hand, we might be tempted to say about acting freely what we saw Graham and Horgan say above about being able to fly. Suppose that some sophisticated metaphysician comes up with a condition, C, that, he purports, is required for a creature to act freely. If we then observe an action that appears to have been performed freely, according to our ordinary standards for such attributions, but fails to meet the putative condition C, what is our reaction? What *should* our reaction be? When it comes to flying ability, our reaction is, and should be, an admission that C must not after all be required for being able to fly. Below I will argue that an analogous reaction is appropriate when it comes to acting freely: the putative condition C must not after all be required for acting freely. In order to make this argument, however, I will need to examine some additional resources from Graham and Horgan (1994).

6.5 Southern fundamentalism as postanalytic metaphilosophy

In later work, Graham and Horgan situate their southern fundamentalism within a broader postanalytic metaphilosophy or methodology (1994). They begin by coining the term "ideology" to refer to the "analysis and clarification of philosophically important ideas or concepts" (1994, 271). Thus, they are proposing a methodology for ideological inquiry—a methodology for what they might call "conceptual analysis," were it not for "the danger that metaphilosophical understanding be skewed by historical connotations of the word 'analysis'." (1994, 290) Their postanalytic methodology can be encapsulated in, and summarized by, the following principles:

(9) The *goal* of philosophy should be to understand how things, in the broadest sense of the term, hang together in the broadest sense of the term.

(10) The proper philosophical methodology is *non-aprioristic*, and does not seek to analyze concepts in terms of noncircular sets of individually necessary and jointly sufficient conditions.

(11) Philosophy should focus on *ideology*, which is "really a broadly empirical, interdisciplinary, enterprise encompassing such fields as psychology, linguistics, social anthropology, and philosophy" (1994, 272).

(12) Even though it is misguided to expect ideological inquiry to produce "high-church" conceptual analyses, the pursuit of *wide reflective equilibrium* among our beliefs and attitudes remains an essential part of philosophy. (This means that philosophy needs to incorporate insights and results from psychology, linguistics, and the sciences.)

(13) Ideological inquirers need not abandon armchair theorizing, because the results of such theorizing can provide defeasible empirical data.

(14) This armchair-obtainable empirical data should set the agenda for subsequent theorizing—in much the same way that the grammaticality judgments of competent language users set the agenda for further linguistic theorizing.

(15) Finally, an adequate account of a concept's ideology should be able to explain "ideological polarity" (1994, 275). Ideological polarity occurs when consideration of a single concept produces intuitions that are incompatible (or at least pull in different directions). A satisfying ideological account needs to explain the pull of whichever intuition(s) it ends up rejecting. This is called the *principle of respect*.

Graham and Horgan's southern fundamentalism is their preferred version of this postanalytic metaphilosophy, and as such has two main components. The first component is a claim about how to resolve the tension in cases of ideological polarity. In such cases, they argue, the tension arises due to a conflict between austere and opulent notions of the concept in question. Moreover, as we briefly saw above, they argue that the austere notion will typically be the one we should accept:

> The key concepts in philosophical problems will normally be relatively austere ideologically; the commitments of statements employing these concepts will normally be no more opulent than is required by the purposes

> for which the concepts are employed in thought, in discourse, and in social practices and institutions. (1994, 280)

This first commitment is called the *principle of ideological austerity*. They argue that familiar empirical data, obtainable from the armchair, support ideological austerity in two ways. (We will look at both of these arguments in more detail below.) First, when we imagine scenarios in which certain features of the opulent concept are missing, we are not inclined to deny the concept's application. For example, consider again *being able to fly*—and suppose that the opulent concept includes a minimum ratio of surface area to body weight. Now imagine a scenario in which some creature appears to be flying, but, we're told, falls short of the minimum ratio. The claim is that we would not (and need not) be inclined to reverse our initial judgment that the creature is flying. Second, they claim, these key concepts will often be such that we can't even conceive of what it would be like to repudiate them (upon coming to the conclusion that one or more conditions for their application aren't met). Given that an austere conception is available, according to which we need not repudiate these concepts even if we discover that the alleged condition(s) for their application aren't met, we should adopt the austere conception (cf. Horgan and Graham 1991, 122–23). These considerations, which can be loosely packaged and labeled as "ideological conservatism," support the principle of ideological austerity.

The second commitment of southern fundamentalism corresponds with (15) above. It is the *principle of respect*, or, a bit more descriptively, the *principle of the respectful explainability of ideological opulence tendencies* (Horgan and Graham 1998, 284ff.). This principle says that the tendency toward opulence—toward endorsing the opulent account of key philosophical concepts—even though usually mistaken, is nevertheless explainable by the nature of the relevant concept and by the cognitive mechanisms of the competent users of that concept. Graham and Horgan propose to explain the intuitive pull of the opulent concepts by appealing to something that we should now be quite familiar with: Lewis's picture of how scorekeeping occurs in a language game. Their Lewisian explanation runs as follows. First, they claim that certain important philosophical concepts contain contextual

parameters—which, recall, means that the standards for evaluating the truth and falsity of claims involving those concepts rise and fall according to conversational context. Second, this context-shifting often occurs unbeknownst to the interlocutors. Third, it's easier for standards to be raised than it is for them to be lowered. Finally, when the standards have been raised, the austere concept will fail to satisfy the raised standards; the opulent concept will have to be employed. This, then, is why the opulent concepts have intuitive pull: it is easy for the relevant standards to be raised to a level such that claims that a particular concept applies will only be true if the concept is taken in the opulent sense—and this shift in standards will often occur without being explicitly recognized by those who are doing the inquiry.

One of Graham and Horgan's examples of ideological polarity (1998, 275) is the concept we are interested in, namely that of acting freely. As perhaps a zeroth approximation, we can explain the relevant opulence tendencies as follows. The ordinary standards for acting freely—for the truth of can-claims—have nothing to do with determinism, which pulls in the direction of compatibilism; but when we explicitly consider the question of whether freedom is compatible with universal causal determinism, there is a noticeable pull towards incompatibilism. According to this postanalytic metaphilosophy, a satisfying ideological inquiry must respect these competing intuitions by explaining the pull of whichever intuition(s) it ends up rejecting. Later in this chapter I will undertake just such an inquiry—which will help us understand examples such as the following, which exemplify precisely the sort of standard-shifting that Graham and Horgan are referring to. The example comes from Galen Strawson (1993, 78), and it involves a thought experiment that

> consists simply in the rigorous application of the belief in determinism to the present course of one's life: one does one's best to think rapidly of every smallest action one performs or movement on makes—or indeed everything whatsoever that happens, so far as one is oneself concerned—as determined; as not, ultimately, determined by oneself; this for a minute or two, say. ... This should have the effect of erasing any sense of the presence of a freely deciding and acting 'I' in one's thoughts; for—so it seems—there is simply no role for such an 'I' to play.

If Graham and Horgan's (1994, 289) Lewisian explanation of the relevant conversational dynamics is right, then we have a case here in which Strawson is not only capitalizing on elevated standards, but *deliberately* elevating them so as to generate the desired intuition. And this is the sort of standard-shifting that occurs when determinism is made salient in a particular conversational context. But before we can offer a contextualist explanation of this kind of example, we need to take a look at some arguments in favor of the first fundamentalist principle: the principle of austerity.

6.6 Arguments for austerity

Recall the epistemic principle that forms the third basic tenet of southern fundamentalism (Horgan and Graham 1991, 109):

> (3) For each PTB condition C, if there were to arise strong epistemic warrant for the thesis that humans do not satisfy C, then (i) this would thereby confer strong epistemic warrant upon the thesis that C is not really a prerequisite for being a true believer, and (ii) it would not confer any significant degree of epistemic warrant upon the thesis that humans are not true believers.

This epistemic principle is clearly offered in the same spirit as the principle of ideological austerity. But what exactly is the relationship between the epistemic principle and the austerity principle? My own view is that the epistemic principle in (3) is plausible on its own, and should be the default methodological position; but to leave it at that would be highly contentious. This is where the austerity principle comes into play. The austerity principle supports and renders even more plausible the epistemic principle: insofar as we have good reason to endorse the austerity principle, we will have good reason to endorse the epistemic principle. And I will argue that the same holds, *mutatis mutandis*, for the epistemic principle in (6): arguments for the austerity of the concept of acting freely will support and render plausible the epistemic principle in (6) . Thus I propose to examine the arguments for austerity.

I begin, as Horgan and Graham do (1991, 117), with the presupposition that the question of whether the concept of *true believer* (for example) is best understood austerely or opulently is an *empirical* question. In other words, the project of figuring out whether a certain concept is austere or opulent is a project whose aim is to provide the best explanation of various empirical facts. Which empirical facts are relevant here? Well, one set of facts includes the judgments of competent language users. These data are brought to bear in the conceptual competence argument.

The conceptual competence argument

As an introduction to the conceptual competence argument, and to better understand in what sense the relevant issues are *empirical* issues, consider first an analogy with natural language syntax. Evaluation of a particular syntactic theory is an empirical project, in the sense that it needs to begin with, or at least be sensitive to, among other things, the grammaticality judgments of competent language users. When there is an intersubjective consensus among the native speakers regarding the grammaticality (or ungrammaticality) of a particular sentence candidate, the best explanation of this consensus is that these competent language users have at least partially latched onto the underlying norms and structures of the language being studied. Given this psychological hypothesis, there is a constraint on proposed syntactic theories—namely that the syntactic judgments of competent language users will turn out to be largely correct (Horgan and Graham 1991, 117). This, then, is the sense in which the search for the best theory of syntax (for a given language or dialect) is an empirical project.

Similarly, the search for the best understanding of true believerhood (and, as we will see, the concept of acting freely) is broadly speaking an empirical project. But the competence that underwrites the austerity of folk psychology is *conceptual* (or semantic), rather than syntactic. Recall the Gettier problem. The story of the Gettier problem in epistemology is the story of a conceptual analysis that was refuted by a class of counterexamples. Certain conditions were proposed, and widely accepted, as necessary and

sufficient for knowledge. But then certain scenarios were proposed in which the subjects met the conditions that were supposed to be sufficient for knowledge, and yet competent language users (in particular, competent epistemologists) consistently and confidently judged that the subjects did not have knowledge of the relevant propositions. The proper response, of course, was not to insist that the subjects had knowledge—contrary to the judgments of competent epistemologists—but rather to acknowledge that the proposed conditions were not after all sufficient for knowledge. And an important reason why this was the appropriate response is that the intuitive judgments of competent epistemologists indicate something about the concept of knowledge. In general, robust patterns of judgment among competent language users "will provide empirical evidence that under an adequate account of the relevant concepts and the terms expressing them, the judgments usually will be correct" (Horgan and Graham 1991, 118).

The next step is to apply this line of reasoning to folk psychology. The facts that need to be explained are the countless ascriptions of beliefs, intentions, and other attitudes, and the wide intersubjective agreement about when such ascriptions are appropriate. Now suppose that we somehow discover that folk psychology is not absorbable into mature science, or that there is after all no language of thought. Would we then stop ascribing beliefs, intentions, and other attitudes? Would there be a radical shift in the standards for when such ascriptions would be appropriate? It seems clear that there would not be any significant pressure to make either of these changes. In other words, these folk psychological judgments are robust, or resilient, with respect to the discovery that various putative true believer conditions don't obtain. (This is, in part, the reasoning behind the epistemic principle in (3).) The best explanation of these robust patterns of judgment is that they indicate conceptual competence among those making the judgments; thus, these patterns provide empirical evidence that an adequate account of the relevant concepts (belief, intention, etc.) will be one in which these judgments turn out to be largely correct. And of the two competing accounts—the austere account and the opulent account—the former is the only one that satisfies this desideratum. Not only that, but the austere account includes

the natural and elegant explanation given above: the *reason why* people consistently and confidently make the folk psychological ascriptions that they do is that such ascriptions reflect their conceptual and semantic competence. In other words, for the most part, people's epistemic standards for these ascriptions are correct: they ascribe beliefs and intentions (etc.) when and only when those ascriptions are appropriate. The opulent account, on the other hand, is one according to which these ascriptions are largely (if not categorically) false, and according to which there is no good explanation for the relevant facts. The opulent account has to somehow explain why people ignore the putative conditions for true believerhood, such as scientific absorbability, and instead use a mistaken epistemic standard that relies on mostly behavioral evidence. For these reasons, the austere account of folk psychology is to be preferred. (Horgan and Graham 1991, 119)

Thus we have good reason—i.e., the conceptual competence argument—to support the austerity principle, and in virtue of that argument we also have good reason to support the fundamentalist epistemic principle in (3).

With a clear grasp of what sort of empirical data we are working with, and the conclusions we can draw from the judgments of (conceptually) competent language users, we are now in a position to provide a conceptual competence argument for the austerity of the concept of acting freely. The relevant facts, which need to be explained, are the countless ascriptions (and assumptions) of free action—the countless number of can-claims—and the wide intersubjective agreement about when such ascriptions are appropriate. There are disputed cases, to be sure, but when it comes to basic actions (e.g., basic bodily movements, such as raising an arm),[72] the patterns of agreement are reliably robust. In general, we are remarkably good at judging actions and, correlatively, evaluating can-claims. Now: suppose

[72] Here I depart from the more common usage of 'basic action' to refer to a *mental* action, such as a decision. On this understanding, a basic action is one that can be performed without performing some other action (cf. Clarke 2003, 18, 80n13). Raising an arm would not be basic in this sense (unless it were completely involuntary, in which case it may not be properly called an "action" at all), but it remains a "basic" and "simple" action in other, less stipulative senses.

that we somehow discover that determinism is true. Fischer (1995, 6) provides an example of what this might look like:

> Suppose ... that a consortium of well-respected scientists announce that they have developed a remarkable new theory which implies that all events can in principle be fully explained by previous events and the laws of nature. That is, they claim that, although they cannot at present make all the predictions about the future, their theory implies that the world is *not* fundamentally indeterministic as many scientists had previously thought; rather, if one knows enough about the past states of the world and the laws of nature, one can confidently predict all the states of the world in the future.

Upon hearing this announcement, how would we, as competent language users, react? Would we then stop ascribing freedom of action to each other? Would we then judge that all can-claims are false? It seems clear that the vast majority of us would not; our freedom ascriptions—our ability judgments—seem resilient with respect to the possible discovery that determinism is true.

This claim is supported by recent experimental work on the question of whether incompatibilism is more intuitive than compatibilism. Nahmias et al. (2006) cite several prominent incompatibilists (e.g., Cover and O'Leary-Hawthorne 1996, Ekstrom 2002, Kane 1999, Pink 2004, G. Strawson 1993) who have claimed, in one way or another, that the folk are typically and pre-theoretically incompatibilists—and hence that incompatibilism is more intuitive than compatibilism. If these incompatibilists are right, that would spell trouble for my thesis; for if most of the folk (i.e., most competent language users) have incompatibilist intuitions, then it is unlikely that their freedom ascriptions would be resilient with respect to the possible discovery that determinism is true. As it turns out, however, it is not at all clear that the folk have intuitive incompatibilist leanings.

Before we look at the details of the experimental work, let us pause to consider whether this question ("Is incompatibilism intuitive?") is even worth asking. I would answer, following Nahmias and co., in the affirmative. While some concepts can perhaps be developed without appealing to ordinary pre-theoretic intuitions, free will is not one of them. Because it is intimately tangled up with our responsibility practices (and with the

reactive attitudes, as we will see below), a theory of free will that does violence to ordinary intuitions incurs a heavy explanatory burden: it must explain both why our intuitions are mistaken and where those mistaken intuitions came from (Nahmias et al. 2006). This is of course reminiscent of the principle of respect, which, as we saw above, is the second major commitment of southern fundamentalism.

In what we might call the fundamentalist spirit, Nahmias and his co-authors give three reasons (2006) why incompatibilists need the support of ordinary intuitions. First, since nothing about determinism logically or conceptually precludes free will, the argumentative burden rests with those who claim that it is impossible for us to act freely if determinism is true. Second, incompatibilism is a more demanding view, metaphysically speaking, than compatibilism. (We might say that the incompatibilist conception of freedom is more "opulent" than the compatibilist's conception.) Thus, it should not be the default position unless it is well-motivated for independent reasons. Third, it is easy to see the motivation for revising our concept of freedom in a way that is metaphysically more benign—even if such a revision brings us into tension with ordinary intuitions. But it is hard to find similar motivation for moving in the opposite direction, metaphysically speaking, if the move toward metaphysical demandingness happens to conflict with ordinary intuitions. So it's important for incompatibilism that it be intuitive—but, unfortunately, it doesn't enjoy the intuitive support that it needs. This in turn provides support for the conceptual competence argument.

Nahmias et al. (2006, 86) take the claim that incompatibilism is intuitive to be equivalent to the claim that the following prediction is true:

> (16) When presented with a deterministic scenario, most people will judge that agents in such a scenario do not act of their own free will and are not morally responsible for their actions.

In order to test this prediction, they ran a study in which they presented participants with various deterministic scenarios. The scenarios differed from each other with respect to how the deterministic nature of the scenario was illustrated, and with respect to the general moral quality (negative, positive, or neutral) of the action.

The first scenario involves a Laplacean conception of determinism:

> Imagine that in the next century we discover all the laws of nature, and we build a supercomputer which can deduce from these laws of nature and from the current state of everything in the world exactly what will be happening in the world at any future time. It can look at everything about the way the world is and predict everything about how it will be with 100% accuracy. Suppose that such a supercomputer existed, and it looks at the state of the universe at a certain time on March 25, 2150 AD, twenty years before Jeremy Hall is born. The computer then deduces from this information and the laws of nature that Jeremy will definitely rob Fidelity Bank at 6:00 pm on January 26, 2195. As always, the supercomputer's prediction is correct; Jeremy robs Fidelity Bank at 6:00 pm on January 26, 2195. (Nahmias et al. 2006, 87)

Participants were presented with this scenario and then asked whether Jeremy acted of his own free will when he robbed the bank. Two variations of this Laplacean scenario (and corresponding question) were also presented to different sets of participants: one that was modified to involve a positive action (saving a child) and one that was modified to involve a neutral action (going jogging). In each case, whether the action was morally negative, positive, or neutral, "a significant majority ... of participants judged that Jeremy does act of his own free will" (Nahmias et al. 2006, 87). And similar results held for moral responsibility—i.e., for the question whether Jeremy was morally blameworthy (for the bank robbery) or morally praiseworthy (for saving the child).

As we saw in the first chapter, it is notoriously difficult to provide a philosophical explication of the thesis of determinism. We should thus expect similar difficulties in presenting the folk with scenarios that attempt to illustrate determinism in a pre-theoretic way. Nahmias and his co-authors acknowledge these difficulties, recognizing the sensitive nature of depicting determinism. On the one hand, it shouldn't be depicted in a question-begging way that involves constraint or coercion; and in fact, given the common juxtaposition of free will and determinism, it's probably best to even avoid use of

"determinism." On the other hand, it should be as salient as possible. So, to ensure the salience of determinism in their study, they develop a second scenario:

> Imagine a universe that is re-created over and over again, starting from the exact same initial conditions and with all the same laws of nature. In this universe the same conditions and the same laws of nature produce the exact same outcomes, so that every single time the universe is re-created, everything must happen the exact same way. For instance, in this universe a person named Jill decides to steal a necklace at a particular time, and every time the universe is re-created, Jill decides to steal the necklace at that time. (Nahmias et al. 2006, 88)

They also develop a third scenario, featuring Fred Jerkson and Barney Kinderson, in which the influence of external causes (genes and upbringing) is made salient through the role that those causes play in bringing about either the stealing or the returning of a wallet containing $1,000. In both of these additional scenarios, judgments about free will and moral responsibility were roughly the same: a significant majority of the participants agreed that the subjects in the scenarios both acted freely and were morally responsible for those actions.

Here is a summary of the results (Nahmias et al. 2006, 89):

Subjects' judgments that the agents …	*Scenario 1 (Jeremy)*	*Scenario 2 (Jill)*	*Scenario 3 (Fred & Barney)*
… acted of their own *free will*	76% (robbing bank) 68% (saving child) 79% (going jogging)	66%	76% (stealing) 76% (returning)
… are *morally responsible* for their action	83% (robbing bank) 88% (saving child)	77%	60% (stealing) 64% (returning)

Table 6.1. Summary of Results

These results clearly suggest that the prediction in (16) is false, and thus that incompatibilism is not after all intuitive. Moreover, these results are applicable to the question at hand—the question of whether ability judgments are resilient with respect to the discovery of determinism. The scenarios developed by Nahmias and his co-authors are scenarios in which

determinism is true; thus the process of presenting participants with these scenarios and asking them questions about the subjects' freedom (or lack thereof) is about as close as we can get to actually determining how competent language users would react upon discovering that determinism is true. I conclude, then, that this study provides us with further reason to accept the conceptual competence argument in favor of conceptual austerity with respect to the notion of acting freely.

As we saw above, one way of clarifying the distinction between the austere conception of folk psychology and the opulent conception of folk psychology was in terms of the alleged gap between resonant intentional systems (RISs) and true believers. According to the austere conception, the gap is minimal: being a non-puppet RIS is sufficient for being a true believer. On the opulent conception, however, there is a significant gap between mere RISs and true believers; various conditions (e.g., the condition that folk psychology be absorbable into a mature science) need to be met in order for a given RIS to count as a true believer. With this in mind, what "gap" can we identify that will help us further distinguish between austere and opulent conceptions of ability? Perhaps the most straightforward way of doing this is to draw a distinction between an *agent* and a *free agent*. However, not all incompatibilists will agree that we are agents, if determinism is true. So it might be better to draw the distinction at the level of actions, and distinguish between an *action* and a *free action*. Almost everyone would agree that my raising my cup of coffee counts as an action, even in a deterministic world; the question, of course, is whether it counts as a *free* action, given the truth of determinism. And the competing answers to this question differ with respect to how much is required, on top of the cup-raising being an action, in order for it to be a free action. Those who favor the opulent notion (i.e., incompatibilists) will require that, at the moment of the raising, the agent raising the cup be able to refrain—consistent with the actual past and the actual laws of nature. Those who favor the austere notion will argue that the gap is metaphysically minimal: the austere concept of acting freely includes fewer conditions than the opulent concept.

Thus we can see how the conceptual competence argument applies to, and supports the austerity of, the concept of acting freely. In virtue of that support, we also have the beginnings of a defense of the fundamentalist epistemic principle in (6):

(6) For each PFA condition C, if there were to arise strong epistemic warrant for the thesis that humans do not satisfy C, then (i) this would thereby confer strong epistemic warrant on the thesis that C is not really a prerequisite for free agency, and (ii) it would not confer any significant degree of epistemic warrant on the thesis that humans are not free agents.

There is, however, more to be said in favor of austerity, and by extension more to be said in favor of (6).

The conceptual conservatism argument

Horgan and Graham (1991) offer another empirical argument for southern fundamentalism (and hence for the austerity principle): the conceptual conservatism argument. As they present this argument, they shift their focus from the concept of belief to the concepts of *action* and *assertion*. The first point about these concepts is that they, like the rest of our language, have evolved to serve certain functions. If we take an etiological view of functions (cf. Wright 1991), then we can go some way toward identifying the function of these concepts by asking what benefit (the use of) these concepts conferred on our ancestors, such that they were reproduced. So, for starters, we can ask about the function of 'action.' What benefit did the use of this concept confer on our ancestors, such that it was reproduced? Well, the concept of action allows us to mark off certain behavior as deliberate, or intentional, or involving practical reasoning and motivational states, rather than a mere movement of the body. This is arguably why we have the concept of action, and hence a good candidate for that concept's function. And we can ask the same question about the concept of assertion: What benefit did the use of the assertion concept confer on our ancestors, such that it was reproduced? This concept apparently allows us to mark off certain

speech acts as passing along information, or transmitting knowledge, or perhaps producing true beliefs in the hearer. (Specifying the exact function of assertion is going to be fraught with controversy. But the point here is merely that there *is* a function, which, when discovered, will explain why we have the concept.)

The next move (the crucial move) in Horgan and Graham's conceptual conservatism argument is the claim that because these concepts (i.e., the concepts of action and assertion) are the products of cultural evolution, and serve certain functions, it's unlikely that they include commitments—conditions on their application—that are not required for the performance of these functions. Thus, if we can plausibly argue that a particular commitment is not required for the performance of a particular concept's function, then we have grounds for dismissing that commitment as a condition on the concept's application. The conceptual conservatism argument, then, provides support not only for the relevant austerity principle but also for the relevant epistemic principle. If we have good conservative grounds for rejecting a particular condition on the application of a particular concept, then we have further justification for rejecting the inference from a failure to meet the condition to the inappropriateness of the concept ascription. In other words, an argument for the superfluity of such a condition supports the recommendation of the fundamentalist epistemic principles encapsulated in (3) and (6). In the case of action (for example), partisans of opulence claim (or at least imply) that the concept includes an implicit commitment to scientific absorbability, or to a language of thought. In other words, according to the opulent conception of folk psychology, the epistemic standards for describing an instance of behavior as an action include requirements such as the requirement that folk psychology be absorbable into a mature science. The concept of action is not properly employed unless scientific absorbability is true. Are either of these conditions required for the performance of the concept's function? Are they required for marking off certain behavior as intentional, or involving motivational states? It's hard to see how a commitment to scientific absorbability, or a language of thought, could contribute to the relevant function. At the very least, there's no natural connection between the question of whether folk psychology is

absorbable into a mature science and the question of whether some agent's action was intentional. (Notice just how remote this absorbability requirement is from the aims and concerns that would lead someone to make, e.g., an action ascription.) It is unlikely, therefore, that the concept of action evolved to include such a requirement, and we (again) seem to be justified in endorsing the fundamentalist epistemic principle in (3).

As above, however, we need to make sure that the conservatism argument for the austerity of folk psychological concepts applies to the concept of acting *freely*. We have seen, in other words, an argument for the claim that various conditions ("putative true actor" conditions, perhaps) are not part of the concept of action; what we now need to show is that a certain putative *free agent* condition is not part of the concept of acting freely. In order to show this, we need to specify the function of the concept of acting freely, and we need to ask whether the relevant condition—in this case, the requirement that we be able to do otherwise, holding fixed the past and the laws—is necessary for the concept's performance of its function.

I actually don't have any firm proposals for the function of the concept of acting freely. Note, however, as we saw earlier in this chapter, that the austere conception of acting freely corresponds roughly to Perry's (2004) "weak account" of ability. And it is a feature of the weak account that we can ascribe an ability to perform an action even if the issue of whether that action will be performed is settled ahead of time. This provides the justification for treating ability ascriptions (can-claims) as roughly synonymous with ascriptions of acting freely, and now gives us a better handle on the relevant function by allowing us to ask about the function of the concept of ability; of 'can.' This question—What is the function of the concept of ability?—is a fascinating one, to which I unfortunately cannot do justice. But I can say a few preliminary things, which will allow us to answer the question that's relevant to the conservatism argument.

First, it's clear that the concept of ability is intimately related to the concept of action, and thus that the function of the ability concept, whatever it may ultimately turn out to be, will most likely involve, to at least some extent, intentions and other hallmarks of

action. If I were to hazard a guess as to the function of our ability concept, I would say that it has survived because of its connection with our desire to bend the world to our ends.[73] We have certain desires, some of which require that something about the world be different. We want to know how likely it is that this change in the world might happen, and so we have developed a concept that indicates a relatively high probability of the occurrence of the desired change. If this is at least a rough approximation of the function of our ability concept, then it seems reasonably clear that being able to do otherwise than we actually do, as an extension of the actual past and consistent with the laws of nature, is no requirement on the deployment of this concept. And even if the actual function of 'ability' is radically different from what I've proposed, it's difficult to imagine how it could be such that the opulent condition is required for the performance of the concept's function. There just doesn't seem to be any reason to think that the "extension of the actual past" condition is going to explain, much less be required for, the evolutionary reproduction of the ability concept.

Implicit in the conservatism argument is the notion that the relevant concepts are *pragmatically indispensable* (Horgan and Graham 1991, 120–23). These notions are pragmatically indispensable because, for better or worse, we cannot give them up. To give up the concept of action would be to renounce action ascriptions; but of course a renouncing is an action. To give up the concept of assertion, at least publicly, would require an assertion. (And so on for most of the other folk psychological concepts, such as belief, intention, epistemic warrant, etc.) In fact, it is difficult to conceive of a scenario in which creatures like us would be able to discard these concepts. The same is obviously not the case with respect to the putative true believer conditions: it is easy to conceive of a scenario in which, e.g., we discovered evidence against, and thus were led to reject, the existence of a language of thought.

[73] This evocative phrase comes from Sosa (2003).

But even if the concept of action is pragmatically indispensable, can the same be said about the concept of acting *freely*? I think it can, and I think the claim that it can is supported by Peter Strawson's (1962) important and influential argument in favor of "optimism" about moral responsibility. Strawson's claim is that even if we were to discover that determinism is true, giving up our responsibility practices is not really a live option for us. Not only do these practices help regulate our behavior, but they also express one of the deepest features of human life: the reactive attitudes. These reactive attitudes—resentment, indignation, gratitude, forgiveness, and love, to name a few—play a crucial constitutive role in our interpersonal relationships. These attitudes give our relationships richness, uniqueness, and distinctiveness. This much is relatively uncontroversial, but Strawson stresses a further point. He argues that because of this crucial interpersonal role, it is clear that a life without the reactive attitudes would not be recognizably human, and hence a practical impossibility. In other words, the reactive attitudes cannot even legitimately be questioned—for these attitudes provide the very framework for questioning their own rationality. To challenge them would be in a sense self-defeating: such a challenge would, practically speaking, undermine its own basis. Thus Strawson: "This commitment [to the reactive attitudes] is part of the general framework of human life, not something that can come up for review as particular cases can come up for review *within* this general framework."[74] (1962, 55, emphasis mine)

Because Strawson's argument primarily concerns moral responsibility (and in particular our responsibility practices, including the reactive attitudes), it's not directly applicable to the question of whether the concept of acting freely is pragmatically indispensable. But if he's right that the various practices associated with moral responsibility (and hence, apparently, the concept itself) are pragmatically indispensable, and if Horgan and Graham are right that the concept of action is pragmatically indispensable, then it would be

[74] This is not the only argument that Strawson puts forth. He also argues that there is no common thread of determinism running between the various cases in which we are liable to excuse people from responsibility, and hence little reason to think that determinism vitiates our responsibility.

odd if the closely related concept of acting freely would differ from its conceptual cousins in this respect. In short, I can't see any reason to think that the concept of acting freely differs from related concepts with respect to pragmatic dispensability.

As we saw above, considerations from Nahmias et al. (2006) lend support to the conceptual competence argument. The same, I submit, can be said with respect to the conceptual conservatism argument. They propose (2006, 96) a corollary of Ockham's razor, which says that "when choosing among theories, all else being equal, we should choose the one that has less metaphysically demanding truth-conditions for its claims." And as it turns out, incompatibilism is more metaphysically demanding: it requires the truth of more metaphysical theses and requires "extra" metaphysical processes (2006, 83). What exactly is a metaphysical process, and what is it to *require extra* metaphysical processes? I'm not sure. But I do think that if we interpret the point in terms of opulence, we can get a handle on what Nahmias and co. are arguing for. Essentially they're arguing for conservatism (and hence austerity) with respect to the concept of acting freely. Whether the point is put in terms of conditions on a concept's application, or the truth of metaphysical theses, or the existence of metaphysical processes, the basic idea is that default view should be one that posits fewer requirements. Philosophical argument might unseat the default, but the starting point should be austere.

Summary

The first principle of southern fundamentalism is that we should prefer austerity. We have seen two arguments for this preference: the conceptual competence argument and the conceptual conservatism argument. An austere understanding of various folk psychological concepts—including, most importantly, the concept of acting freely—provides us with the best explanations for the following data: the robust and wide-ranging intersubjective judgments of competent language users, and the fact that we can easily conceive of abandoning various putative true believer conditions, but can't conceive of abandoning the core concepts of folk psychology. (In addition, the austere conception is preferable to the

extent that it coincides with the apparent function of the relevant concepts.) With austerity thus defended, we are now in a position to examine the second southern fundamentalist principle: the respectful explainability of ideological opulence tendencies. Before we do so, however, and as preparation for what's to come, I would like to revisit and emphasize a point from the previous chapter (as briefly alluded to in the introduction to this chapter): Lewis's (1996) claim that we have no reason to accept the primacy of contexts in which standards of attribution (of, e.g., knowledge) have been raised as a result of conversational mechanisms. In other words, claims Lewis, we have no reason to believe that the knowledge that is correctly attributed to a subject in an ordinary standards case is in any way inferior to the knowledge that is correctly attributed in a raised-standards context.

What we see now, in light of these arguments for austerity, is that we not only have negative reasons for resisting the primacy of high(er)-standards contexts; we also have *positive* reasons for sometimes preferring the ordinary standards context. In particular, if the way in which the standards of some context have been raised involves the implicit imposition of the sort of condition that plays no role in the explanation of the evolution of the relevant concept, and if that condition is such that attributions of the concept wouldn't be abandoned even if it were established that the condition didn't obtain, then we have reason to prefer the context in which ordinary standards are in force.

Thus ends our discussion of the first southern fundamentalist principle (aka the austerity principle). But before we move on to the second principle, we need to examine a few objections to the fundamentalist project.

6.7 Objections to fundamentalism

Horgan and Graham (1991, 123–29) consider and respond to several objections to their folk psychological realism, two of which are particularly relevant to free will fundamentalism. The first objection can be put in question form: What exactly is the fundamentalist's stance toward the allegedly opulent conditions on acting freely, such as the requirement that the

agent be able to do otherwise as an extension of the actual past (consistent with the laws), and how does this mesh with the epistemic principle in (6)? (Recall (6):

> (6) For each PFA condition C, if there were to arise strong epistemic warrant for the thesis that humans do not satisfy C, then (i) this would thereby confer strong epistemic warrant on the thesis that C is not really a prerequisite for free agency, and (ii) it would not confer any significant degree of epistemic warrant on the thesis that humans are not free agents.

My fundamentalist response to this objection, following Horgan and Graham, is that these PFA conditions are like engineering hypotheses; they are at best *de facto* prerequisites, as opposed to conditions that are dictated by the very concept of a free action. The analogy of flying ability is once again helpful: if being able to fly does require a certain minimum ratio of body surface to body weight, then that requirement is a *de facto* requirement, rather than a condition built directly into the concept of flying ability. Similarly, if acting freely does require that we enjoy a certain relationship to the past and the laws, then that requirement is a *de facto* requirement—rather than a condition dictated by the very concept of acting freely. And what (6) tells us is that if we were to discover that that requirement is not met in our case, then the right (i.e., the warranted) thing to do is to conclude that the condition is not after all a prerequisite for acting freely.

The next objection asks: Why we don't just say the following? Acting freely requires being able to do otherwise, consistent with the nomologically actual past, and since it's very likely that we sometimes act freely it's also very likely that we have the ability to do otherwise, consistent with the actual past and laws. The answer is that we shouldn't say that, because if we were to discover evidence that determinism is true, then we would thereby discover evidence that we never act freely. And at that point we would be faced with a choice: either give up on the idea that we sometimes act freely (an idea that enjoys an extremely high degree of antecedent warrant—recall fundamentalist tenet (5), above), or give up on the idea that being able to do otherwise, consistent with the actual past and laws, is

required for acting freely. (Giving up on the condition would be akin to what Fischer and Ravizza (1998, 253) have called "metaphysical flip-flopping.") If, on the other hand, we endorse fundamentalism, then we will not be faced with such a difficult choice.[75] We will instead be able to maintain the thesis that we act freely while explaining why we need not maintain the doctrine that acting freely requires being able to do otherwise as an extension of the actual past.

This point can be put in general terms, as a *resiliency* constraint on conceptual analysis (or "ideological inquiry," to use Graham and Horgan's preferred nomenclature). This constraint says that one's conceptual analysis should not be determining the answer to an empirically open question. One should be careful, in other words, about dictating empirical results from one's armchair. (Or, conversely, the claim is that we should be careful lest our conceptual analysis be undermined by empirical discoveries.) The constraint can be stated slightly more precisely (and narrowly) as follows:

(17) A conceptual analysis should not imply the falsity of an empirical thesis that is epistemically open.

[75] The difficulty here arises because of the need to give up one of two commitments, both of which are quite plausible. If they were equally plausible, then the fundamentalist claim that we need to give up on one, rather than other, might be difficult to motivate. But notice that the plausibility enjoyed by the first commitment (i.e., our commitment to sometimes acting freely) is a different kind of plausibility than the plausibility enjoyed by our second commitment (i.e., our commitment to being able to do otherwise, consistent with the actual past and laws, as part of the concept of free action). The plausibility of the first commitment comes from various sources—including the arguments for austerity that we looked at above. The second commitment is intuitively plausible, and supported by the consequence argument, but its plausibility is (at least in my view) not as robust. In addition, the fact that its plausibility can be explained, compatibly with it being false, without needing to impute radical and wide-ranging error to competent language users, makes it a better candidate for denial than the first commitment (the plausibility of which cannot be explained compatibly with its falsity without imputing radical and wide-ranging error to competent language users.)

Although I can't offer an ironclad defense of this constraint, it does strike me as intuitively plausible, and I will provide what I take to be a strong motivation for endorsing it. I will begin by emphasizing an undesirable consequence of failing to satisfy it.

Imagine that the question of whether some empirical claim is true or false is epistemically open, and that some conceptual analysis entails that the empirical claim is false. Imagine further that scientists discover irrefutable evidence that the empirical claim is true. In this scenario, the conceptual analysis has been proven inadequate, and therefore must be abandoned. Now imagine a different analysis (of the same concept), one that does not entail the falsity of the epistemically open empirical claim. Even if scientists were to prove the truth of the empirical claim, proponents of this second analysis would not be forced to give up the analysis. All other things being equal, it seems that in this scenario the second conceptual analysis (i.e., the one that does not entail the falsity of the empirical claim) is to be preferred to the first. In other words, given a conceptual analysis that entails the falsity of an epistemically open empirical claim, there is always the possibility that a competing analysis will capture the relevant phenomena equally well without dictating the empirical result—and hence will be superior to the first analysis. Given this possibility, it seems to be a constraint on a conceptual analysis that it not come with this *built-in* possibility of a better analysis—i.e., that it not entail the falsity of an epistemically open empirical claim.

The claim, then, is that, other things being equal, conceptual analysis A is to be preferred to conceptual analysis B if A is neutral with respect to an epistemically open empirical claim the falsity of which is entailed by B. In this scenario, we can describe analysis A as more resilient (with respect to empirical discovery) than analysis B; hence, I am proposing a *resiliency constraint* on conceptual analysis.

This is not to say that this particular consideration necessarily trumps any of the other virtues of conceptual analysis. Instead, I am merely suggesting that if two analyses capture the relevant phenomena equally well, and if one of the analyses is more resilient with respect to empirical discoveries, then the resilient analysis is to be preferred. In other words,

I am suggesting that this proposed resiliency constraint carries roughly the same weight as the *simplicity* constraint that we often use in evaluating scientific theories. Briefly, the simplicity constraint says that if two theories explain the data equally well, then the one that is simpler (e.g., contains fewer laws, or fewer variables, or posits fewer unobservables) is to be preferred. Thus, the resiliency constraint resembles the simplicity constraint in this respect: it is clearly not inviolable, but violating it should be considered a cost. It may be a cost worth paying on occasion, but it remains a cost nonetheless.

Isn't this just backtracking compatibilism?

Another objection that might be raised against my view runs as follows: If the austere conception of ability is to be construed, as above, as a weak account of ability (in Perry's (2004) terms), then isn't this view just another version of backtracking compatibilism (and hence vulnerable to the objections discussed in Chapter 2)? In response, notice that the austere concept we are working with is defined negatively: it is the conception of ability such that a certain opulent condition doesn't apply (or at least doesn't fall out of the conception itself). And since that opulent condition imposes the requirement that I be able to do something other than what I actually do, given the actual past and the actual laws of nature, there are two ways of rejecting that condition. We can allow that if I were to do otherwise, the past would have been different, or we can allow that if I were to do otherwise, an actual law of nature would not have been a law of nature. (There is of course a third option as well, namely that if I were to do otherwise, both the past and the laws would have been different.) The fundamentalist contextualist compatibilist need not take a stand on which of those two options is preferable. If both of them turn out to be hopeless, then things would be different; but given the state of the dialectic, as represented in Chapters 2 and 3, I see no reason to be so pessimistic about their ultimate prospects.

Putting the "mentalism" in fundamentalism

The final objection is that free will fundamentalism is a form of *mentalism,* and therefore unsatisfying as a philosophical analysis. The distinction here is between mentalism and *extra-*mentalism, and it's described by Goldman and Pust (1998) as follows:

> Broadly speaking, views about philosophical analysis may be divided into those that take the targets of such analysis to be in-the-head psychological entities versus outside-the-head non-psychological entities. We shall call the first type of position *mentalism* and the second *extra-mentalism*. ... In the case of mentalism, the mental entities in question are not conscious mental entities; otherwise, it would presumably be unnecessary to use indirect inferential techniques to get at them. Rather, they are non-conscious entities or structures to which introspective access is lacking. (Goldman and Pust 1998, 183–84)

With this distinction in mind, the objection is that while fundamentalism does give us some valuable insight about what's going on inside the heads of competent language users, it tells us nothing about non-psychological entities. Goldman and Pust explain why this is alleged fact about fundamentalism is supposed to be problematic in terms of Markie's (1996) complaint about a mentalist approach to epistemic justification:

> Markie claims that the mentalist account of our judgments of justification does not offer any explanation of what *makes* a belief justified or a process justification conferring. Instead, it offers us only a psychological hypothesis about how people make epistemic judgments, about how they come to classify beliefs into epistemological categories. For this reason, Markie claims that the mentalist version of the reliability theory of justification is perfectly consistent with positions usually thought opposed to reliabilism. (Goldman and Pust 1998, 195)

The complaint, in short, is that mentalist accounts of epistemic justification might tell us when a belief deserves to be *called* justified, but they do not tell us what justification *is*. Goldman and Pust offer a two-fold response to this complaint. Their first point is that they do in fact offer an explanation, and even an account, of what makes a belief justified:

> The mentalist claims that what makes a belief justified is possession of those properties by virtue of which a deployer of the justification concept, who is aware of all relevant properties of the target belief, would classify it as justified. An account of what those properties are is the very thing that a developed mentalist proposal will provide. (1998, 195–96)

Their second point is that the complaint seems to be motivated by some version of the extra-mentalist proposal—which faces problems of its own. Although I don't wish to wade into the broader fray, and take a stand on whether mentalism or extra-mentalism should be the preferred form of philosophical analysis, I do think it's worth attempting to respond to a Markie-type complaint against free will fundamentalism.

The complaint, as it relates to the concept of acting freely, is that the fundamentalist account of our freedom judgments (or, roughly synonymously, our judgments with respect to ability claims) does not offer any explanation of what *makes* an action free; it does not tell us what free action *is*. In response to the complaint, I think we can first of all help ourselves to a variation on what Goldman and Pust offer. Even if the fundamentalist proposal is a mentalist proposal, the claim is going to be that what makes an action free is the possession of those properties by virtue of which a deployer of the freedom concept (or the ability concept), who is aware of all relevant properties of the target action, would classify it as free. An account of what those properties are is the very thing that a developed (funda)mentalist proposal will provide. It is true that I have not provided such an account (at least not in any great detail). But such an account is not necessary for my purposes, which is to provide a response to the consequence argument. (Although it is, of course, a natural choice for the next item on the agenda.) All that's needed is a likelihood of such an account, and I see no reason to think that the prospects are hopeless.[76]

6.8 Contextualism

Moving now to the second fundamentalist principle (aka the principle of respect), notice that it is something of a self-corrective: it says that a preference for austerity should be

[76] John Maier's (2008) defense of "epistemic compatibilism" suggests an even stronger answer to this objection, namely the response that the extra-mentalist demands too much: "The best we can expect of a defense of a position like compatibilism is a principled account of how we ought to *apportion our credences,* given our imperfect epistemic situation." (2008, 90, emphasis in original)

coupled with an explanation of why there might be an intuitive pull toward opulence. As a requirement on proponents of an austere conception, this seems uncontroversial. Significantly more controversial—and where our dispute takes place—is the question of whether the proffered explanation is a good one. As we saw above, Horgan and Graham provide, following Lewis, a brief contextualist explanation for the tendency toward ideological opulence. I would like to develop this explanation in detail, as a way of fleshing out the fundamentalist view that I have been proposing. (Defending this principle, in other words, is where we start putting the contextualism in fundamentalist contextualist compatibilism.)

The first step in my development of free will fundamentalism involved arguing for the primacy of the austere conception of acting freely. The second step involves providing an explanation of the intuitive pull of the opulent conception. Only with such an explanation can we be fully justified in endorsing free will fundamentalism. I begin by considering a recent attempt (Hawthorne 2001) to develop a contextualist version of compatibilism, and some of the objections to it. Responding to these objections will allow me to develop my own implementation of contextualism about acting freely—one that will explain the intuitive pull of incompatibilism while avoiding the pitfalls of Hawthorne's version.

Hawthorne's contextualism

We begin with Hawthorne (2001), who was one of the first to lay out a contextualist picture of freedom. His (2001) is a development of a suggestion made in O'Leary-Hawthorne and Pettit (1996), which is built on Lewis's (1996) contextualism about knowledge. Lewis, recall, tries to dodge what he considers to be a false choice—between skepticism and fallibilism—by analyzing knowledge that *p* in terms of eliminating possibilities that not-*p* (1996, 694):

(18) *S knows* that *p* iff *S*'s evidence eliminates every possibility in which not-*p*—Psst!—except for those possibilities that we are properly ignoring.

The contextualist nature of this analysis, as we saw above, has to do with the rules that Lewis specifies as governing whether an ignored possibility is properly ignored—rules that will vary according to conversational context.

Hawthorne (2001) develops a parallel analysis of acting freely, according to which the relevant question is not whether one's evidence eliminates a certain possibility, but instead whether one's action is free from "causal explainers" beyond one's control. (A causal explainer, as Hawthorne understands it, is a state of affairs that figures into an adequate causal explanation of some event. I will follow him in taking this term to be roughly synonymous with "causal factor," "causal influence," and "causal determinant.") Thus Hawthorne (2001, 68):

> (19) *S* does *x* freely only if *S*'s action is free from causal explainers beyond *S*'s control—Psst!—apart from those causal explainers that we are properly ignoring.[77]

As with Lewis's account of knowledge, the contextualist nature of Hawthorne's account of freedom has to do with which ignored explainers are *properly* ignored—a fact that will vary according to conversational context. Hawthorne doesn't attempt to develop these rules, but he does consider three strategies for such a development: a consequentialist strategy, a descriptive strategy, and a transcendental strategy. The consequentialist strategy (2001, 71) says, roughly, that causal influences are properly ignored if ignoring them will produce more overall happiness than not ignoring them would. The descriptive strategy (2001, 72) looks at actual human practices and attempts to systematize these folk practices into rules for proper ignoring. The transcendental strategy—which originates with Strawson (1962)—"proceeds by considering which kinds of ignorings constitute the conditions under which the participant stance is possible (whereby human reactive attitudes like gratitude and resentment are

[77] Hawthorne (2001, 68) also emphasizes the deliberate choice of "only if" rather than "if": "What I have to say does not require that the right hand side is sufficient as well as necessary. A good thing too, since the stronger 'iff' claim is far more questionable."

possible)." (Hawthorne 2001, 72) Hawthorne doesn't take a position on which of these strategies is most promising, although he does disparage (in a footnote) the descriptive strategy:

> I myself have little sympathy with this broad approach. Thinking that one can analyze the rules of propriety for attending to causal forces by doing descriptive anthropology about when folk are disposed to raise their eyebrows is rather akin to thinking that one can settle debates in legal philosophy concerning when mental disorders abnegate accountability simply by investigating the dispositions of ordinary folk to attribute or deny accountability when informed of the disorder. (Hawthorne 2001, 79n10)

Although I will not pursue the descriptive strategy at any great length, I think Hawthorne's dismissal of it is premature. This is because the analogy he draws is not apt. Recall the conceptual competence argument in favor of austerity, according to which the best explanation for wide intersubjective agreement about, e.g., freedom ascriptions is that those making the ascriptions enjoy (and share) a certain competence with the relevant concepts. If that argument goes through, then we could indeed learn something about which causal influences are properly ignored by examining when the folk are "disposed to raise their eyebrows." But this is not analogous to looking at when the folk are disposed to attribute or deny legal accountability in cases involving mental disorder. First, I doubt that there is a similarly robust intersubjective agreement about when attributions of legal accountability are appropriate. Of course, that's an empirical question, and I could be wrong about it. But at the very least it seems clear that agreement about legal responsibility attributions is not *as* robust, not *as* widespread, as is the agreement about freedom attributions. Moreover, there is a handy explanation (of this lack of agreement) readily available: even though we are typically competent with respect to attributions of *moral* accountability in normal cases, competence with respect to attributions of legal accountability requires additional conceptual tools, and additional knowledge, relative to what the average person possesses. And since the appropriateness of legal accountability attributions gets even trickier when mental disorders are involved, even more tools and knowledge are required. It's not hard to see how the relevant competencies could diverge as we move from ordinary freedom

attributions to the trickier attributions of legal accountability, and insofar as these competencies might diverge we have reason for rejecting Hawthorne's analogy.

Whatever the ultimate set of rules of propriety ends up looking like, it will, according to Hawthorne, include the *rule of attention*: the rule stating that if a causal explainer is attended to, then it's not properly ignorable. This rule is an essential part of Hawthorne's treatment of the consequence argument. As we saw in Chapter 1, the consequence argument relies on a transfer principle connecting *p* (which is a complete description of some past state of the world, including the laws of nature) and *q* (which is an action performed by *S,* and is entailed by *p*):

(20) If *S* has no power over whether *p,* and no power over whether *p* entails *q,* then S has no power over whether *q*.

The key to the contextualist response to the consequence argument, according to Hawthorne, is to apply the rule of attention. When we're wondering whether *S* has done *x* freely, we're asking whether there are any causal explainers of *x* that are outside of *S*'s control. But of course the answer is "Yes" when we're considering the consequence argument. When we're considering the consequence argument (more specifically, the transfer principle), we're in essence asking ourselves if S has power over whether *q,* given that *p* entails *q* (which follows from the assumed truth of determinism), that she doesn't have power over *p,* and that she doesn't have power over the fact that *p* entails *q*. As Hawthorne points out,

> It is certainly true that when presented with any instance of van Inwagen's principle, we find it compelling. Present us with a determining cause over which an agent has no control and a law that connects the cause to that agent's doing *x,* and we find ourselves obliged to say the agent did not do *x* freely. But notice that in considering any particular instance of the principle, we ipso facto *attend* to a postulated cause. Exploiting this fact, the contextualist can explain the appeal of each instance of van Inwagen's principle that we imaginatively consider without endorsing the principle itself. (2001, 73–74, emphasis in original)

Notice here that what the contextualist is able to do is *explain the appeal* of the consequence argument. The consequence argument highlights, and thus forces us to attend to, causal factors over which we have no control. The rule of attention tells us that we cannot properly ignore these causal factors, once they're attended to, and as a result we will never be able to attribute freedom to an action to which we have applied the consequence argument.

Hawthorne actually goes farther than this (2001, 74), and suggests endorsing the following transfer principle instead:

(21) For all *p* and *q:* If *S* has no power over whether *p* (and *p* is not properly ignored), and if *S* has no power over whether *p* entails *q* (and that entailment is not properly ignored), then *S* has no power over whether *q*.

Thus, according to Hawthorne, van Inwagen's principle, as represented in (20) is false—because there are some causes that are properly ignorable. (21) is superior because it represents an acknowledgment of this fact, and because it explains why the consequence argument is so compelling.

We have been looking for an explanation of the intuitive force of incompatibilism, and now we have one. From thinking about how Lewis's contextualism might apply in the present case, we have learned that the plausibility of incompatibilism—which depends largely on the plausibility of the consequence argument—is due to a conversational mechanism that raises the standards for attribution of the concept of acting freely. And with a little help from Hawthorne, we learn the details of how this works. The relevant standard for freedom attributions pertains to which causal determinants can be properly ignored: a higher standard restricts the number of causal determinants that can be properly ignored. And the very terms in which the consequence argument is formulated are terms according to which the ultimate causal determinant(s) cannot be ignored; so it's no surprise that upon consideration of the consequence argument, we will be inclined to deny that the action in question is performed freely.

As is to be expected, there are some objections to Hawthorne's view, which need to be considered before we can move forward. (Hawthorne himself is not likely to consider such objections, as he basically disavows the account in the end (2001, 77): "For convenience of exposition, I have represented myself as confidently believing a Lewis-style contextualist analysis of freedom. In fact I do not confidently believe it. It is a controversial philosophical thesis and I don't confidently believe many of those." This view is thus somewhat of an orphan. I will try to briefly provide it with a home.)

Objections to Hawthorne's contextualism

The first objection comes from Steven Rieber (2006), who has also presented a contextualist account of freedom, and is thus largely in sympathy with Hawthorne. His primary complaint is that Hawthorne's account is *ad hoc* and poorly motivated:

> The major drawback to Hawthorne's account is that the contextualism is unmotivated apart from its capacity to solve the puzzle. That is, no independent reason has been given for thinking that ascriptions of freedom are context-sensitive in the strong sense required by the contextualist. (Rieber 2006, 230)

Rieber thus attempts to find independent motivation for being a contextualist about free will. He analyzes an agent's acting freely in terms of the agent being the *original cause* of the action, and then develops and defends two rules for evaluating the truth of claims involving the notion of "original cause." Contextualism about free will follows from these two rules, and is thus independently plausible to the extent that Rieber's rules are.

In response, I should say first that I don't really feel the force of Rieber's complaint. If there is an interesting puzzle, and contextualism is capable of solving that puzzle, then that seems to be motivation enough for developing a contextualist solution. But if special motivation is needed, then it has arisen over the course of my development of fundamentalism about free will. We have seen arguments for austerity with respect to the notion of acting freely, and we have also seen the need to provide an explanation of the

temptation toward opulence. The contextualist move is designed to meet this need for an explanation, and thus seems motivated insofar as the search for such an explanation.

The next objection (actually a set of objections) comes from Richard Feldman (2004). (I will consider the first two here and the third later in the chapter.) He claims, first, that contextualism about both knowledge and freedom are *prima facie* implausible. He contrasts those concepts with the concept of hunger (2004, 266). If I say that I'm hungry, and then eat a meal, and then say that I'm not hungry, it would not concern me if someone pointed out the apparent contradiction between my two statements. My lack of hunger now gives me no reason to doubt my earlier claim to hunger. Feldman's point is that this is not how, e.g., freedom ascriptions work. If freedom ascriptions worked in this way, then if some subject S were to begin doubting her freedom on the basis of the consequence argument—perhaps even to the point of saying that she's not free—then she would not be inclined to see that claim as contradicting earlier (pre-consequence-argument) claims to freedom. But, claims Feldman, this is not what happens. Instead, S would see her denial of freedom as in conflict with earlier attributions.

This objection, it seems to me, is just another version of the semantic blindness objection, which was addressed in Chapter 5. (Recall that the semantic blindness objection points to [and attempts to problematize] the contextualist admission that we may not be aware of contextual shifts, and as a result may not notice some of the consequences of context-sensitivity.) It is perhaps a bit surprising that we would mistakenly assume that certain freedom (or knowledge) ascriptions conflict—but, as it turns out, there are other examples of semantic blindness (e.g., flatness) and the invariantist, as we saw in Chapter 5, has his own version of the problem of semantic blindness. Absent some additional reasons to think that semantic blindness creates problems for contextualism in general, we can safely set aside this first objection.

Feldman's second objection is that contextualism about freedom concedes too much to the incompatibilist:

> If philosophical settings make the standards such that the sentences ascribing 'free' are not true, then in philosophical settings such as this one, the incompatibilists (and those who deny that we are free) are right. The philosophical debate, which obviously occurs in philosophical contexts, is won by the incompatibilists. Indeed, the compatibilists lose without much of an argument in Hawthorne's account. They capitulate from the outset. (2004, 272)

This objection goes through if we start, as Hawthorne does, in neutral territory. But I have gone to some length to defend an *austere* conception of freedom as the proper starting point. The fundamentalist move is what motivates and supports a compatibilist position; it is in fact the opposite of capitulation. Contextualism, I will readily admit, cannot get one all the way to compatibilism; but it can provide a plausible explanation of why the consequence argument drives us into the cold arms of incompatibilism.

6.9 Fundamentalist contextualist compatibilism

With my proposed versions of fundamentalism and contextualism explicated and defended, we are now in a position to see how my proposal—fundamentalist contextualist compatibilism—serves as a response to the consequence argument. Recall once more the structure of the argument:

(20) If *S* has no power over whether *p,* and no power over whether *p* entails *q,* then *S* has no power over whether *q*. (Transfer of powerlessness)

(22) *S* has no power over whether the past and the laws entail that she perform some action *X*. (Premise)

So, (23) If *S* has no power over the past or the laws, then S has no power over whether she does *X*. (20, 22)

(24) But *S* has no power over the past or the laws. (Premise)

So, (25) *S* has no power over whether she does *X*. (23, 24)

If determinism is true, we can suppose that the past and the laws entail that, for example, *S* will raise her coffee cup to take a drink. Thus, according to the consequence argument, just prior to her raising of the cup it will be false that

(26) *S* can refrain from raising her cup.

According to the free will fundamentalist, (26) is clearly and obviously true (assuming normal conditions and no funny business involving agential constraint, cercion, or manipulation) according to the ordinary behavior-based standards for the truth of can-claims. If the fundamentalist is right about this, then he will have to deny (25) as well, and hence (23) or (24). The claim in (24) seems difficult to deny, so that leaves the fundamentalist with (23). And what grounds might we have for denying (23)? Well, notice that (23) represents the opulent condition that I have been at pains to reject as required for the truth of can-claims. (23) represents the notion that acting freely (i.e., having power over an action) requires the ability to perform or refrain from the action as an extension of the actual past. In fact, (23) follows from this notion, together with the assumption that determinism is true. If acting freely requires extending the nomologically actual past, and if determinism is true, then acting freely requires being able to change the past or the laws. But we have seen why this opulent notion of free action is the wrong notion, and is conceptually inferior to the austere notion. And the contextualist move provides the explanation of why, when presented with the consequence argument, we are inclined to give the opulent notion more weight than it deserves. Hawthorne's (2001) explanation of the appeal of the consequence argument, which we briefly examined above, is apt: When we consider the fact that, given the truth of determinism, our actions are entailed by the past and the laws of nature, then we are inclined to think that those actions are not performed freely. But of course from their being entailed by the past and the laws of nature it doesn't follow that they are restricted in any of the ways that lead us in ordinary contexts to deny freedom of action. Thus, in order to judge that our actions aren't free, we need to impose an additional opulent condition on acting freely—a

condition that's inconsistent with our actions being entailed by the past and the laws of nature. Fundamentalist contextualist compatibilism is the view that this imposition, while understandable, is not warranted.

6.10 Conclusion

In this chapter I have presented and developed what I call "fundamentalist contextualist compatibilism." I would like to close by briefly emphasizing how this approach to free will is superior to the "merely" contextualist approach championed by, e.g., Hawthorne (2001).

Recall that one of Feldman's (2004) criticisms of Hawthorne's view is that, relative to other forms of compatibilism, it concedes too much to the incompatibilist. This concession is explicit in the following passage from Hawthorne:

> Note, then, that I concede a great deal more to the skeptic that the standard compatibilist. Having done my philosophy, I am quite prepared to say with the philosopher who denies freedom "People don't act freely." I am quite unwilling to say that such a philosopher is abusing the English word "free" or replacing it with a new concept when she says "People don't act freely." (Hawthorne 2001, 69)

For Feldman, this concession amounts to capitulation. But the fundamentalist contextualist compatibilist can do better. He understands why some philosophers say that "People don't act freely"—and he understands why such philosophers can convince others to say the same—but he does not himself see any good reason to say this. And notice that the fundamentalist is not doing what Hawthorne finds objectionable: the fundamentalist is not claiming that incompatibilists are abusing the English word "free," or replacing it with a new concept. What incompatibilists are doing instead is something more subtle: they are using the transfer of powerlessness principle (or similar considerations) to raise the standards for acting freely—to add a requirement for acting freely that is typically not understood to be in force—and thereby inclining themselves and others to deny that we ever act freely (if determinism is true).

We are now in a position to consider a final objection from Feldman (2004). Feldman argues that Hawthorne's contextualist analysis, because it includes only a necessary condition for acting freely, fails to address some of the central concerns that an account of acting freely should address. What Feldman would like to see, and what he thinks an analysis of freedom needs, is a *sufficient* condition for acting freely. In other words, Feldman (2004, 273) claims that a satisfying analysis of freedom needs to at least go some way toward answering certain questions, chief among them the following: "is the (or a) sense of the word 'free' such that it can apply to actions even though they are determined?" I will agree with Feldman that Hawthorne's account may not go far enough on this score. But here again I think the fundamentalist preference for austerity, as well as the epistemic principle that dictates our response upon discovering that certain putative conditions for acting freely are not met, provides an answer to Feldman's main question. There is indeed a sense of the word 'free' that can apply even to determined actions, and that sense is the austere sense.

The lesson here is that Hawthorne's account is (as he himself would acknowledge) incomplete; it's not able to bear the weight that it needs to bear in order to be a fully satisfying analysis of acting freely. But if it is employed as I am employing it—as an explanation of incompatibilist tendencies given the superiority of the austere conception of freedom—then the incompleteness objection does not apply.[78] Thus fundamentalist contextualist compatibilism provides a satisfying response to the consequence argument without falling prey to some of the objections that apply to a mere contextualist compatibilism.

[78] Willaschek (forthcoming) argues that Feldman's objections apply to Rieber's (2006) view to the same extent that they apply to Hawthorne's view, and that they force a different approach to contextualism about free will. According to Willascheck's analysis (forthcoming, 5), *S* performs *A* freely just in case (1) she is able to (a) form a considered practical judgment about whether or not do A and (b) act in accordance with that judgment; and (2) she is the original cause of her doing *A*. Willaschek (forthcoming, 16) then analyzes "original causation" so that "the concept of free will is context-sensitive in a way that does not make ascriptions of free will vary with context." He considers his view to be superior to Hawthorne's (2001) and Rieber's (2006) because it is a contextualist compatibilism that doesn't fall prey to Feldman's objections. I welcome him (or perhaps he will welcome me) to the club.

• Chapter 7 •

The epistemological treatment

7.1 Introduction

In this concluding chapter I will summarize the progress we've made and draw out some connections at the big-picture level. I will also explore some possibilities for future research at the intersection of agency theory and epistemology.

7.2 Recapitulation

We began, following Fischer (1995), by noticing and emphasizing a parallel between two arguments: the consequence argument (which purports to establish that acting freely is incompatible with causal determinism) and the skeptical argument in epistemology (which purports to establish that we know very little, if anything, about the external world). One of the salient similarities between these two arguments is that they both rely on a closure, or *transfer* principle. The consequence argument relies on a transfer of powerlessness principle and the skeptical argument relies on a transfer of knowledge principle (which is more commonly referred to in the epistemological literature as the epistemic closure principle). The similarities between these two arguments led us to ask whether a response to one can be modified and applied to the other.

In preparation for that undertaking, we examined two influential responses to the consequence argument: backtracking compatibilism and local miracle compatibilism. We saw that backtracking compatibilism is notable for its commitment to the claim that we can do otherwise than we actually do, even if determinism is true, and that if we were to do otherwise, then the past would have been different than it actually was. Following Perry (2004), we distinguished between a *strong* account of ability and a *weak* account of ability.

The strong account supports, and is characterized by, the claim that if it is *settled* ahead of time that some agent *S* will perform (or refrain from) some action *A,* then *S* is not able to refrain from (or perform) *A*. The weak account, on the other hand rejects this commitment, and in so doing tries to separate the question of whether it is settled that *S* perform *A* from the question of whether *S* can refrain from performing *A*. The motivation behind this separation is the conviction that the correct theory of ability will not include conditions that refer to what does the settling, namely the entire past history of the world together with the laws of nature. (Compare "financial ability": the correct theory of financial ability will refer to financial resources, and not to the mental states of the one considering making a purchase. It may be settled that the purchase not be made, but this doesn't preclude the financial ability to make the purchase.) Despite the plausibility of this view, it faces at least two problems. First, it seems to lead to some counterintuitive results in the sphere of practical reasoning. Second, it's not clear that the distinction between a proposition's being *made true* (which precludes an agent's doing something such that it would have been false) and its being settled (which does *not* preclude an agent's doing something such that it would have been false) can do the work that the backtracking compatibilist needs it to do. While these objections may not be devastating, they do motivate an examination of an alternative response to the consequence argument: local miracle compatibilism.

Local miracle compatibilism is notable for a commitment of its own—namely its commitment to the claim that we can do otherwise, even if determinism is true, and if we were to do otherwise then an actual law of nature would not have been a law. This commitment leads proponents of the local miracle view to endorse the following relatively weak nomological principle as encapsulating the sense in which our actions are constrained by the laws of nature: We cannot do anything that would *be* or *cause* an event that violates the laws of nature. We noted first of all (in Chapter 2) that whereas one way to be a local miracle compatibilist is to endorse a weak theory of the laws (according to which events *establish* the laws by confirming them [or failing to disconfirm them]) rather than a strong theory of the laws (according to which events *conform* to the laws), this is not the only way to

be a local miracle compatibilist. In fact, the proponent of the local miracle view need not take a firm stand on which theory of the laws is correct. We then examined (in Chapter 3) an influential criticism of local miracle compatibilism (Ginet 1990), which claims that the nomological principle endorsed by the local miracle compatibilist is unable to explain our *inabilities* in a wide range of cases. In response, I pointed out that the principle in question need not explain our inability in these cases, because it is the compatibilist's *theory* of ability that is going to do the necessary explanatory work. Again it becomes clear that questions about can-claims (e.g., about whether I can refrain from raising my cup) are typically—and properly—answered by looking to a theory of ability rather than looking to the question of whether causal determinism obtains.

As a brief aside, I will attempt to strengthen this point a bit by expanding upon some considerations from Chapter 3. Let us begin with something that seems to be obvious, namely that there are two related questions here, the answers to which can come apart. First question: Can I raise my cup in a deterministic world? (Of course.) Second question: Given that I raise my cup in a deterministic world, could I have refrained from raising my cup? (Maybe not. It's hard to say.) And it seems to me that the answer to the first question is going to depend on one's analysis of *can,* whereas the answer to the second question is going to depend on other things (e.g., the success or failure of the consequence argument). If this is right, then there are two different notions in play here. The first notion is simply the notion of ability—of 'can.' The second notion is the notion of being able to do otherwise, holding fixed the past and the laws. And these two notions, being different, require (or at least allow for) two different analyses. (The difference between these two notions can be obscured by the fact that sometimes when we ask whether someone can do something, what we're really asking is whether she could have done otherwise, given that she did something else.)

Notice the implications of denying this distinction. To deny the difference between these notions is to say that indeterminism is a condition on being able to do things (e.g., my being able to raise my cup). But surely indeterminism isn't a condition on my being able to

do things. (One could, I suppose, argue that indeterminism is always an implicit condition on being able to do things—but it's hard to see any advantage to saying that there's one concept with a deeply implicit condition, rather than just saying that there are two concepts with different conditions.)

Or consider another way of putting the point, in terms of the distinction between the 'can' of general ability and the "all-in" 'can' (cf. Fischer 2008), where the all-in 'can' is the more particularized concept that is linked to moral responsibility as the necessary and sufficient freedom-relevant condition. One way to construe the compatibility question, I take it, is as the question of whether the concept of 'can' that's relevant to free will is the all-in concept or the general concept. (Another and perhaps better way to construe the question would be as the question of whether the all-in can should be interpreted in a libertarian way, or in a compatibilist way.) What I'm suggesting—what I think is suggested by the above treatments of backtracking compatibilism and local miracle compatibilism—is that however the compatibility question is construed (and whether or not my own answer to that question is the correct one), when we're *analyzing* 'can,' we should be dealing with the general concept rather than the all-in concept.

Returning to the main line of argument, we can see that the insights gleaned from backtracking compatibilism and local miracle compatibilism take us some of the way, but certainly not all of the way, toward the desired response to the consequence argument. Further progress requires taking a step back from the consequence argument and examining (in Chapter 4) the skeptical argument in epistemology. Inspired by our initial treatment of the consequence argument, we considered a preemptive response to the skeptical argument, namely that it begs the question. What we saw was that the skeptical argument does presuppose an absolutist conception of knowledge according to which all possibilities of error must be ruled out. But while it may be difficult to justify this absolutist conception on independent grounds, it is also difficult to come up with a principled way of drawing a line between possibilities that need to be ruled out and possibilities that can be safely ignored. We considered two principles—the *sensitivity* principle and the *safety* principle—that

purported to provide the distinction we need, but we saw that both principles (and especially the sensitivity principle) pressure us to deny epistemic closure. So we reopened the question (first broached in Chapter 1) of whether to accept the epistemic closure principle, and considered several examples (from Hawthorne (2004), who was himself building on the discussion in Vogel (1999)) designed to establish the falsity of epistemic closure. In the end, though, we determined that the costs of denying closure are too great, and thus that we should be wary of endorsing principles that incline us toward such a denial. This conclusion prompted the search for a response to the skeptical argument that did not require us to reject the absolutist notion of knowledge (and hence adopt a principle, such as sensitivity, that is in tension with closure).

We found such a response (in Chapter 5) by appealing to Lewis's (1979a, 1996) contextualism. Lewis posits a conversational mechanism that governs and explains potential shifts in the boundary between possibilities that can be ignored and possibilities that cannot be ignored. The basic idea is that these shifts generally occur in a way that will allow statements made in a conversation to be evaluated as true (or at least conversationally acceptable). If, for example, additional possibilities need to be relevant in order for a particular claim to be true, then those possibilities will typically become relevant. Lewis's account provides several insights. First, it is easier to raise the relevant standards than it is to lower them once they have been raised. Second, the fact that a claim is false relative to raised standards does not imply that it was false relative to earlier, unraised standards. Third, there is no reason to think that contexts in which the standards are raised are in any way superior to original contexts in which the standards remain unraised. And finally, one of the virtues of Lewis's account is that it rejects the skeptical argument while providing an *explanation* of why the argument can seem so compelling. This will be taken as a desideratum of a satisfying response to the consequence argument.

In Chapter 6 we bring together the various pieces, developed in earlier chapters, that together constitute my proposed response to the consequence argument. But before we can build the response, we need to lay the foundation. This foundation is a modified version of

southern fundamentalism (Graham and Horgan 1991, 1994), which is a philosophical methodology that seeks to adjudicate between *austere* and *opulent* notions of important philosophical concepts. An austere conception of *acting freely* (for example) is one according to which the conditions for the truth of can-claims are fairly minimal: they correspond roughly to the behavior-based epistemic standards evinced by competent language users. An opulent conception of acting freely, on the other hand, posits additional conditions for the truth of can-claims. For example, the opulent conception is one according to which an agent cannot act freely unless she can do otherwise than she actually does, holding fixed the past history of the actual world and the actual laws of nature. The free will fundamentalist endorses an epistemic principle according to which reasons to believe that opulent conditions do not obtain are not reasons to deny the *application* of the relevant concept, but instead reasons to deny that *those conditions are required* for the application of that concept. When we apply this methodology to the concept of ability (to do otherwise than we actually do), what we get looks a lot like contextualism about ability claims. On this view, the concept that is most relevant to mundane usage is the austere concept—according to which freedom to do otherwise is not ruled out by causal determination. (In other words, according to the austere concept, it is not a condition on my acting freely that I be able to do otherwise than I actually do, holding fixed the actual past and the laws of nature.) However, when philosophers begin to investigate the concept, the opulent notion often comes into play. And according to the opulent notion, freedom is incompatible with determinism. This provides us with a way of affirming freedom, even in the face of determinism, while explaining how we can be led (astray) into thinking that the truth of determinism would preclude our freedom. (It also, incidentally, provides us with some additional support for the idea, introduced in Chapters 2 and 3, that a correct theory of ability will not include conditions involving the past or the laws.)

Thus we can characterize fundamentalist contextualist compatibilism according to two main tenets: First, there are general empirical considerations that support the austere (compatibilist) concept of freedom. Second, we can explain the persistence of

incompatibilist intuitions by pointing out that the ordinary standards for what counts as freedom are easily raised—which is why the opulent concept often mistakenly seems to be appropriate. In particular, we can see that the consequence argument exerts its force by raising the standards for acting freely, such that it requires being able to do otherwise as an extension of the actual past (consistent with the laws of nature). Moreover, the attempt to separate questions about can-claims from questions about determinism (as discussed in Chapters 2 and 3) is buttressed by the fundamentalist commitment to the primacy of the austere conception of acting freely.

This, then, is fundamentalist contextualist compatibilism: an empirically-based commitment to the austere conception of acting freely, together with a contextualist explanation of the apparent force of the consequence argument. This commitment, when coupled with its attendant explanation, constitutes a powerful new defense of compatibilism about freedom and determinism.

7.3 Extension

I would now like to consider some avenues for future research. To the extent that one is convinced, as I am, that insights from epistemology can be fruitfully incorporated into the dialectic(s) surrounding free will and moral responsibility, one will have reason to seek out other parallels between epistemology and agency theory. I think there are several obvious parallels that can be drawn, and I would like to briefly discuss them here. But first, by way of motivation for this discussion, recall again Lewis's (1979a) proposal for the rule of accommodation that governs *relative modality*. Ordinary language modal verbs, such as *can,* and *knows,* are not absolute; in other words, various possibilities can be ignored when we're evaluating the truth of can-claims or knowledge-claims. This deep similarity between these modal verbs further motivates the search for fruitful parallels between the fields of study that have grown up around them.

I think there are several key concepts in agency theory that could benefit from what we might call the epistemological treatment. This treatment, at least as I am conceiving of

it, involves construing threats to various aspects of our agency as running in parallel to the skeptical argument in epistemology. There are different ways of responding to the skeptical challenge in epistemology, and I think some of these ways can be adopted and adapted for use in agency theory. We have seen how this works with respect to the concept of acting freely, but I would like to briefly sketch out some ways in which anti-skeptical insights from epistemology might shed light on some of the other key agency-related concepts: in particular, moral responsibility and our responsibility practices of praise, blame, and punishment.

The "impossibility" challenge

Even if our actions are free in the sense that we are able to do otherwise, there is a further question as to whether we are morally responsible for those actions. This challenge to our agency comes in the form of an argument that such moral responsibility is *impossible*. For it might seem that in order to be responsible for the actions I take, I have to be responsible for the *source* from which those actions flow. In other words, I have to be responsible for the way I am: my character, my motives, and so on. But the only way I can influence my character is through my actions—which themselves flow from my character. And so I seem to be stuck in a circle, with no chance of finding room for responsibility. In short, the argument is this: moral responsibility for actions requires *ultimate responsibility* (for character), but ultimate responsibility is impossible. So moral responsibility for actions is impossible (cf. G. Strawson 1986). Since this argument applies with equal force to everyone, it threatens our sense of the moral quality of our actions because it suggests that moral responsibility itself is impossible. Let us call this the "impossibility challenge."

We saw in Chapter 4 that one common and influential response to the skeptical challenge is the *fallibilist* response. We also saw that fallibilism, although initially plausible, is not without its costs—including a commitment to the truth of "clashing conjunctions" (Fantl and McGrath 2009) such as

(1) I know that *p* but it's possible that not-*p*.

Whether or not these concerns are decisive against the fallibilist, there may be some room for a "fallibilist" approach to moral responsibility. Drawing inspiration from Feldman (2003), we might say the following. First of all, there are linguistic reasons to deny the need for ultimate responsibility. Consider the following conjunction:

(2) I'm responsible for what I did, even though I'm not *ultimately responsible* for the entirety of my character.

If ultimate responsibility (for character) were required for moral responsibility for actions, then an assertion of (2) should sound paradoxical, or at least produce some discomfort. To my mind, however, there is no such discomfort. (Far from it, in fact: (2) seems to be quite sensible, the sort of thing I would remind myself of were I attempting to misguidedly absolve myself of blame for some action of mine.)

There are also practical reasons to deny the need for ultimate responsibility. For even if Galen Strawson (1986) and his impossibilist sympathizers are right, and responsibility for actions is impossible, there is still an important concept—call it responsibility*—that we would need to invent. For we need a way to distinguish between those actions over which someone has absolutely no control (e.g., actions that are the result of addiction, or the result of coercion) and those actions which flow from the mechanism of a properly functioning agent who is responsive to reasons—i.e., actions that we typically classify as actions for which agents are responsible. Moreover, given the obvious richness and importance that would attach to our theorizing about responsibility*—and the fact that the set of actions for which agents are *ultimately* morally responsible would be (and is) empty—it would be no big loss to give up the pursuit of, and inquiry into, ultimate responsibility. In light of all this, it seems best to just to drop the ultimacy criterion from the concept of responsibility.

I have called this last point in favor of optimism a practical point; but that locution is perhaps misleading. For while this "practical" defense of fallibilism about responsibility is certainly sensitive to practical concerns, it stands or falls on conceptual grounds. The impossibilist has pointed out that there is a certain conception of responsibility—ultimate responsibility—that is impossible; he also notes that such an ultimacy criterion is inconsistent with our ordinary normative commerce. In light of this inconsistency (i.e., in light of the fact that this conception of responsibility is unattainable in practice), the impossibilist proposes that we revise our practices accordingly. But the fallibilist can (and should) point out that there is another, and apparently superior, alternative when one becomes aware of the conflict between our responsibility practices and the ultimacy criterion: revise the conception of ultimate responsibility with which our practices conflict. (This move, it should be clear by now, shares certain affinities with southern fundamentalism.)

Fischer (2006) makes a similar point. He is also addressing Galen Strawson's argument, and in particular the notion of self-creation that underlies Strawson's conception of moral responsibility. In light of the evident conflict between our ordinary responsibility practices and the purported need for self-creation, Fischer asks (2006, 110): "Why ... is it more plausible to jettison moral responsibility and cling to a very demanding notion of self-creation ... than to scale down the demands of self- creation to something more reasonable?" Instead of insisting on this radical conception of responsibility, it seems more reasonable to "suppose that we must be the 'ultimate sources' of our behavior in some genuine way, but a way that is at least possible to realize in the world" (Fischer 2006, 112). Operating under this supposition has the benefit, among others, of preserving the richness of our theorizing without forcing us to create an entirely new field of inquiry—one that studies precisely what we thought we were studying all along.

The challenge from luck

Even if some of our actions are free, and it's at least possible that we are morally responsible for such actions, there's a concern that threatens to undermine some of the *practices* associated with moral responsibility: in particular the practices of praise, blame, and punishment. This is the problem of *moral luck,* which can be stated as follows. Most of us think of ourselves as generally decent persons. For example, I think that I am generally a decent person, and that is because most of my actions are under my control—and of those actions under my control, none of them (at least so far as I know) have been morally heinous. Moreover, when I am praised for what seem to be praiseworthy actions, and blamed for what seem to be blameworthy actions, I often find those reactions appropriate. In general, we typically evaluate ourselves and others on the basis of what we (and they) have control over, and these evaluations typically form the basis for our judgments of praiseworthiness and blameworthiness. But there are some cases in which the intuitive moral assessment of two agents varies significantly on the basis of factors that are outside of the agents' control. A classic example, of course, is the drunk driver (cf. Nagel 1976). Consider two drunk drivers, Cliff and Floyd. They are equally drunk, and thus equally culpable for choosing to drive. Suppose that Cliff makes it home safely, while Floyd, unfortunately, hits and kills a pedestrian. Intuitively, what Floyd did, and what he is thus punished for, is much worse than what Cliff did. And yet the difference between what they did was entirely out of their control. (The point could be extended to cover overall assessments of their lives as well: a life that includes killing a pedestrian is, other things being equal, worse than a life that does not.)

Exacerbating this problem are some recent findings from social psychology, which indicate that seemingly insignificant *situational* factors can have a dramatic impact on our behavior (cf. Doris 2002). To pick one disturbing example, the best explanation of why so many ordinary German citizens participated in the mass killings of the Holocaust appears to be a situationist one: it's not that all or even most Germans in the 1940s were murderously anti-Semitic (*contra* Goldhagen 1997); rather, imperceptible but powerful social pressures produced shockingly high levels of compliance in the groups (e.g., the police battalions)

responsible for many of the killings. If this is right, then it seems that the difference between me, an ordinary American citizen, and an ordinary German citizen who participated in the mass killings of the Holocaust has more to do with our respective circumstances or situations than it has to do with our respective characters. Were I in his situation, the chances are good that I would have done many of the same horrific things. The upshot is that it seems to be a matter of luck that I am (mildly) praiseworthy for being a morally decent person—an unsettling conclusion to say the least.

I think we can go some way toward answering this challenge by attending to some recent work from Pritchard (2005, 2006). He draws a distinction between different kinds of epistemic luck, and argues that it can be applied to questions about moral luck—helping us understand the various ways in which it affects the moral quality of our actions. The basic idea is that there are two kinds of epistemic luck: *veritic luck* and *reflective luck*. When a belief is subject to veritic luck, it's a matter of luck that the agent's belief is true. (This, for example, is the kind of luck that afflicts the subjects of the Gettier cases.) Reflective luck is a bit different: we say (or at least Pritchard says) that a belief exhibits reflective luck when, *given only what the agent is able to know by reflection alone,* it is a matter of luck that her belief is true. The distinction can be brought out by considering the case of the "chicken-sexer"—an example that's widely used in the debate between internalists and externalists about epistemic justification. Imagine someone who has a natural and reliable ability to distinguish between male and female chicks—but she doesn't know how she does it, and in fact isn't even aware that she has this ability. Call this individual the "naïve chicken-sexer." The question is whether the naïve chicken-sexer can know that a particular chick is, say, male. Externalists will typically allow that she might have such knowledge, while internalists will typically deny knowledge. The "enlightened" chicken-sexer, on the other hand, is aware of her ability, and has good reasons for believing it to be a reliable ability, and even has some understanding of how it works. Whether or not the naïve chicken-sexer knows that the chick is male, it does seem as though the enlightened chicken-sexer is in a better epistemic position. At any rate, we can easily imagine a point along the chicken-sexer spectrum (from

naïve to enlightened) at which the chicken-sexer is still subject to some degree of reflective epistemic luck and yet knows the sex of a particular chick. Thus, it seems that *some* degree of luck is compatible with knowledge.

To the extent that we can plausibly draw a parallel between these two challenges from luck (i.e., the challenge to our responsibility practices and the challenge to our knowledge), we can suggest that some degree of luck is compatible with praising, blaming, and punishing. It will of course, take some work to establish exactly where we should draw the line between the amount of luck that does not undermine our responsibility practices and the amount of luck that does. But there is hope for this project if we consider the intriguing possibility (proposed in Pritchard 2005) that there is actually no problem of moral luck—that there is only a problem of epistemic luck. According to this proposal, the problem of moral luck can be reduced to the problem of reflective epistemic luck. If this is right (although unfortunately I can't render a verdict here), then the solution to the problem of moral luck stands or falls with the solution to the problem of reflective epistemic luck.

7.4 Conclusion

Traditional defenders of our agency have marshaled the resources of metaphysics, ethics, and the philosophy of mind, with mixed results, in the ongoing attempt to explain how we can be free and responsible creatures in the face of various threats. (See, for example, Doris 2002, Franklin 2010, Libet 2001, Nahmias et al. 2008, Nichols and Knobe 2008, Skyrms 2004, and Walter 2001.) However, despite the persistent and increasingly sophisticated efforts of these philosophers, the perennial threats persist, and new ones arise. Since I count myself among those attempting to defend our agency, I applaud these recent efforts. But in the midst of this understandable enthusiasm for interdisciplinary dialogue, there has been an unfortunate neglect of *intra*disciplinary dialogue: there have been very few attempts to consult areas of philosophy other than metaphysics, ethics, and the philosophy of mind. More specifically, there is one particular area of philosophy—epistemology—that has been sorely neglected by most of those who theorize about free will and moral responsibility. This territory at the

intersection of agency theory and epistemology strikes me as woefully under-explored, and the present work represents an attempt to remedy that deficiency in some small way.

References

Audi, R. 1978. "Avoidability and Possible Worlds." *Philosophical Studies* 33: 413–21.

Beebee, H. 2000. "The Non-Governing Conception of Laws of Nature." *Philosophy and Phenomenological Research* 56: 571–94.

Beebee, H. and Mele, A. 2002. "Humean Compatibilism." *Mind* 111: 201–23.

Bennett, J. 1984. "Counterfactuals and Temporal Direction." *Philosophical Review* 93: 57–91.

Brueckner, A. 1985. "Skepticism and Epistemic Closure." *Philosophical Topics* 13: 89–117.

Campbell, J. K., O'Rourke, M., and Shier, D. (eds.). 2004. *Freedom and Determinism.* Cambridge, MA: MIT Press.

Cappelen, H. and Lepore, E. 2005. *Insensitive Semantics.* Malden, MA: Blackwell Publishing.

Cartwright, N. 1999. *The Dappled World.* Cambridge: Cambridge University Press.

Ciulla, J. B., Martin, C., and Solomon, R. C. (eds.). 2007. *Honest Work: A Business Ethics Reader.* New York: Oxford University Press.

Clarke, R. 2003. *Libertarian Accounts of Free Will*. New York: Oxford University Press.

Cohen, S. 1998. "Contextualist Solutions to Epistemological Problems: Skepticism, Gettier, and the Lottery." *Australasian Journal of Philosophy* 76: 289–306. Reprinted in Sosa, et al. (eds.), 706–20. Page references are to the reprinted version.

———. 2004. "Knowledge, Assertion, and Practical Reason." *Philosophical Issues* 14: 482–91.

———. 2010. "Stewart Cohen." In Dancy, Sosa, and Steup (eds.), 118–23.

Cover, J. A. and O'Leary-Hawthorne, J. 1996. "Free Agency and Materialism." In *Faith, Freedom, and Rationality,* J. Jordan and D. Howard-Snyder (eds.), 47–71. Lanham, MD: Rowman and Littlefield.

Dancy, J., Sosa, E., and Steup, M. (eds.). 2010. *A Companion to Epistemology, 2nd edition*. Malden, MA: Blackwell Publishing.

Dennett, D. 1984. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, MA: MIT Press.

DePaul, M. R. and Ramsey, W. (eds.). 1998. *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*. Lanham, MD: Rowman & Littlefield.

DeRose, K. 1995. "Solving the Skeptical Problem." *Philosophical Review* 104: 1–52. Reprinted in Sosa, et al. (eds.), 669–90. Page references are to the reprinted version.

———. 1999. "Contextualism: An Explanation and Defense." In J. Greco and E. Sosa (eds.), *The Blackwell Guide to Epistemology,* 187–205. Malden, MA: Blackwell Publishing.

———. 2002. "Assertion, Knowledge, and Context." *Philosophical Review* 111: 167–203.

———. 2005. "The Ordinary Language Basis for Contextualism and the New Invariantism." *Philosophical Quarterly* 55: 172–98.

———. 2009. *The Case for Contextualism*. New York: Oxford University Press.

Doris, J. 2002. *Lack of Character*. New York: Cambridge University Press.

Dretske, F. I. 1970. "Epistemic Operators." *The Journal of Philosophy* 67: 1007–23.

———. 1971. "Conclusive Reasons." *Australasian Journal of Philosophy* 49: 1–22.

Earman, J. 1986. *A Primer on Determinism*. Dordrecht: D. Reidel.

Ekstrom, L. W. 2002. "Libertarianism and Frankfurt-Style Cases." In Kane (ed.), 309–22.

Fantl, J. and McGrath, M. 2009. *Knowledge in an Uncertain World*. New York: Oxford University Press.

Feldman, Richard. 2003. *Epistemology*. Upper Saddle River, NJ: Prentice Hall.

———. 2004. "Freedom and Contextualism." In Campbell, O'Rourke, and Shier (eds.), 255–76.

———. 2007. "Knowledge and Lotteries." *Philosophy and Phenomenological Research* 75: 211–26.

Fischer, J. M. 1979. "Lehrer's New Move: 'Can' in Theory and Practice." *Theoria* 45: 49–62.

———. 1995. *The Metaphysics of Free Will*. Malden, MA: Blackwell Publishing.

———. 2006. "The Cards That Are Dealt You." *Journal of Ethics* 10: 107–29.

———. 2007. "Compatibilism." In J. M. Fischer, R. Kane, D. Pereboom, and M. Vargas, *Four Views on Free Will,* 44–84. Malden, MA: Blackwell Publishing.

———. 2008. "My Way and Life's Highway: Replies to Steward, Smilansky, and Perry." *Journal of Ethics* 12: 167–89.

Fischer, J. M. and Ravizza, M. (eds.). 1993. *Perspectives on Moral Responsibility*. Ithaca, NY: Cornell University Press.

———. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press.

Franklin, C. E. 2010. *Libertarianism, Freedom, and the Brain* (PhD dissertation: University of California, Riverside).

Gettier, E. 1963. "Is Justified True Belief Knowledge?" *Analysis* 23: 121–23.

Ginet, C. 1966. "Might We Have No Choice?" In K. Lehrer (ed.), *Freedom and Determinism,* 87–104. New York: Random House.

———. 1980. "The Conditional Analysis of Freedom." In P. van Inwagen (ed.), *Time and Cause: Essays Presented to Richard Taylor,* 171–86. Dordrecht: D. Reidel.

———. 1983. "In Defense of Incompatibilism." *Philosophical Studies* 44: 391–400.

———. 1990. *On Action*. New York: Cambridge University Press.

Goldhagen, D. J. 1997. *Hitler's Willing Executioners: Ordinary Germans and the Holocaust.* New York: Vintage.

Goldman, A. and Pust, J. 1998. "Philosophical Theory and Intuitional Evidence." In DePaul and Ramsey (eds.), 179–97.

Graham, G. and Horgan, T. 1994. "Southern Fundamentalism and the End of Philosophy." *Truth and Rationality: Philosophical Issues* 5: 219–47. Reprinted in DePaul and Ramsey (eds.), 271–92. Page references are to the reprinted version.

Graham, P. A. 2008. "A Defense of Local Miracle Compatibilism. *Philosophical Studies* 140: 65–82.

Graham, P. J. 2006. "Liberal Fundamentalism and Its Rivals." In J. Lackey and E. Sosa (eds.), *The Epistemology of Testimony,* 93–115. New York: Oxford University Press.

———. 2007. "The Theoretical Diagnosis of Skepticism." *Synthese* 158: 19–39.

Hawthorne, J. 2001. "Freedom in Context." *Philosophical Studies* 104: 63–79.

———. 2004. *Knowledge and Lotteries*. New York: Oxford University Press.

———. 2006. "Testing for Context-Dependence." *Philosophy and Phenomenological Research* 73: 443–50.

Hoefer, C. 2010. "Causal Determinism." In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2010 Edition). URL = <http://plato.stanford.edu/archives/spr2010/entries/determinism-causal/>.

Horgan, T. 1977. "Lehrer on 'Could'-Statements." *Philosophical Studies* 32: 403–11.

———. 1979. " 'Could,' Possible Worlds, and Moral Responsibility." *Southern Journal of Philosophy* 17: 345–58.

Horgan, T. and Graham, G. 1991. "In Defense of Southern Fundamentalism." *Philosophical Studies* 62: 107–34.

Ichikawa, J. 2009. "Keith DeRose, *The Case for Contextualism*." In *Notre Dame Philosophical Reviews.* URL = <http://ndpr.nd.edu/review.cfm?id=18307>.

Kane, R. 1999. "Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism." *Journal of Philosophy* 96: 217–40.

———. (ed.). 2001. *The Oxford Handbook of Free Will*. New York: Oxford University Press.

Knobe, J., and Nichols, S. (Eds.). 2008. *Experimental Philosophy*. New York: Oxford University Press.

Kvanvig, J. 2004. "Nozickian Epistemology and the Value of Knowledge." *Philosophical Issues* 14: 201–18.

Lehrer, K. 1976. " 'Can' in Theory and Practice: A Possible Worlds Analysis." *Action Theory: Proceedings of the Winnipeg Conference on Human Action,* 241–70. Reprinted in John Martin Fischer (ed.), *Free Will: Critical Concepts in Philosophy*, Vol. IV, 234–61. New York: Routledge. Page references are to the reprinted version.

Lewis, D. 1979a. "Scorekeeping in a Language Game." *Journal of Philosophical Logic* 8: 339–59.

———. 1979b. "Counterfactual Dependence and Time's Arrow." *Noûs* 13: 455–76. Reprinted in Lewis (1986), 32–66.

———. 1981. "Are We Free to Break the Laws?" *Theoria* 47: 113–21. Reprinted in L. W. Ekstrom (ed.), *Agency and Responsibility: Essays on the Metaphysics of Freedom*, 30–37. Boulder, CO: Westview Press, 2001. Page references are to the reprinted version.

———. 1986. *Philosophical Papers: Volume II*. New York: Oxford University Press.

———. 1996. "Elusive Knowledge." *Australasian Journal of Philosophy* 74: 549–67. Reprinted in Sosa, et al. (eds.), 691–705. Page references are to the reprinted version.

Libet, B. 2001. "Do We Have Free Will?" In Kane (ed.), 551–64.

Loewer, B. 1996. "Humean Supervenience." *Philosophical Topics* 24: 101–27.

Maier, J. 2008. *The Possibility of Freedom* (PhD dissertation: Princeton University).

Markie, P. 1996. "Goldman's New Reliabilism." *Philosophy and Phenomenological Research* 56: 799–817.

Nagel, T. 1976. "Moral Luck." Reprinted in T. Nagel, *Mortal Questions,* 24–38. New York: Cambridge University Press, 1979.

Nahmias, E., Morris, S. G., Nadelhoffer, T., and Turner, J. 2006. "Is Incompatibilism Intuitive?" *Philosophy and Phenomenological Research* 73: 28–53. Reprinted in Knobe and Nichols (eds.), 81–104. Page references are to the reprinted version.

Nichols, S. and Knobe, J. 2007. "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions." Reprinted in Knobe and Nichols (eds.), 105–26.

Nozick, R. 1981. *Philosophical Explanations*. Cambridge, MA: Harvard University Press.

O'Leary-Hawthorne, J. and Pettit, P. 1996. "Strategies for Free Will Compatibilists." *Analysis* 56: 191–201.

Perry, J. 2004. "Compatibilist Options." In Campbell, O'Rourke, and Shier (eds.), 231–304.

———. 2008. "Can't We All Just Be Compatibilists? A Critical Study of John Martin Fischer's *My Way*." *The Journal of Ethics* 12: 157–66.

Pink, T. 2004. *Free Will: A Very Short Introduction*. New York: Oxford University Press.

Pritchard, D. H. 2002. "Resurrecting the Moorean Response to Skepticism." *International Journal of Philosophical Studies* 10: 283–307.

———. 2005. *Epistemic Luck.* New York: Oxford University Press.

———. 2006. "Moral and Epistemic Luck." *Metaphilosophy* 37: 1–25.

Rieber, S. 2006. "Free Will and Contextualism." *Philosophical Studies* 129: 223–52.

Skyrms, B. 2004. *The Stag Hunt and the Evolution of Social Structure.* New York: Cambridge University Press.

Slote, M. 1982. "Selective Necessity and the Free-Will Problem." *The Journal of Philosophy* 79: 5–24.

Sosa, E. 1999. "How to Defeat Opposition to Moore." *Philosophical Perspectives* 13: 141–54. Reprinted in Sosa, et al. (eds.), 280–89.

———. 2000. "Skepticism and Contextualism." *Philosophical Issues* 10: 1–18.

———. 2002. "Tracking, Competence, and Knowledge." In P. K. Moser (ed.), *The Oxford Handbook of Epistemology,* 264–87. New York: Oxford University Press.

———. 2003. "The Place of Truth in Epistemology." In M. DePaul and L. Zagzebski (eds.), *Intellectual Virtue: Perspectives from Ethics and Epistemology,* 155–79. New York: Oxford University Press. Reprinted in Sosa, et al. (eds.), 477–91.

———. 2004. "Replies." In J. Greco (ed.), *Ernest Sosa and His Critics,* 275–326. Malden, MA: Blackwell Publishing.

Sosa, E., Kim, J., Fantl, J., and McGrath, M. (eds.). 2008. *Epistemology: An Anthology*. 2nd edn. Malden, MA: Blackwell Publishing.

Stack, M. K. 2010. "For Poland, plane crash in Russia rips open old wounds." *Los Angeles Times*, April 11.

Strawson, G. 1993. "On 'Freedom and Resentment'." In Fischer and Ravizza (eds.), 67–100.

Strawson, P. 1962. "Freedom and Resentment." *Proceedings of the British Academy* 48: 1–25. Reprinted in Fischer and Ravizza (eds.), 45–66. Page references are to the reprinted version.

Unger, P. 1979. *Ignorance: A Case for Scepticism*. New York: Oxford University Press.

van Inwagen, P. 1974. "A Formal Approach to the Problem of Free Will and Determinism." *Theoria* 40: 9–22.

———. 1975. "The Incompatibility of Free Will and Determinism." *Philosophical Studies* 27: 185–99.

———. 1977. "Reply to Narveson." *Philosophical Studies* 31: 89–98.

———. 1983. *An Essay on Free Will*. Oxford: Clarendon Press.

———. 2004. "Freedom to Break the Laws." *Midwest Studies in Philosophy* 28: 334–50.

Vihvelin, K. 2000. "Libertarian Compatibilism." *Philosophical Perspectives* 14: 139–66.

Vogel, J. 1990. "Are There Counterexamples to the Closure Principle?" In M. D. Roth and G. Ross (eds.), *Doubting*, 13–27. Dordrecht: Kluwer Academic Publishers. Reprinted in Sosa, et al. (eds.), 290–301. Page references are to the reprinted version.

———. 1999. "The New Relevant Alternatives Theory." *Philosophical Perspectives* 13: 155–80.

Walter, H. 2001. "Neurophilosophy of Free Will." In Kane (ed.), 565–76.

Warfield, T. A. 2004. "When Epistemic Closure Does and Does Not Fail: A Lesson from the History of Epistemology." *Analysis* 64: 35–41.

Watson, G. (ed.). 2003. *Free Will*, 2nd ed. New York: Oxford University Press.

Wiggins, D. 2003. "Towards a Reasonable Libertarianism." In Watson (ed.), 94–121.

Willaschek, M. Forthcoming. "Non-Relativist Contextualism About Free Will." *European Journal of Philosophy*.

Williamson, T. 2000. *Knowledge and Its Limits.* New York: Oxford University Press.

Wright, L. 1999. "Functions." In D. J. Buller (ed.), *Function, Selection, and Design*, 29–56. New York: SUNY Press.