

Agentive Experience Compatibilism

by

Oisín Deery

B.A. Philosophy and English, The National University of Ireland, Galway, 1993

M.A. Philosophy, University College, Cork, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE STUDIES

(Philosophy)

The University Of British Columbia

(Vancouver)

August 2013

© Oisín Deery, 2013

Abstract

Libertarians, who think that freedom is incompatible with determinism and we are free, claim that their view is descriptively right about how we view ourselves as agents who are free to do otherwise. Much of what compatibilists, who think that freedom is compatible with determinism, have written in recent years about the freedom to do otherwise has consisted in attempts to deflate these sorts of libertarian claims. Philosophers on each side thereby make claims about the nature of our experience and beliefs. These are empirical claims, which can be illuminated by empirical methods. In experiments that I ran with Matt Bedke and Shaun Nichols, our participants described their experience of being free to do otherwise as incompatibilist across a range of conditions. Compatibilists may dismiss these results by insisting that people are mistaken in their introspective judgments. Or they might insist that people only seem to have incompatibilist beliefs about freedom and determinism; in fact, these beliefs are about the compatibility of freedom and fatalism. I argue that both these compatibilist claims are false, at least in the forms that they currently take. Instead, I argue that compatibilists should concede that people's experience and beliefs are in part libertarian, but can still be accurate if determinism is true. First, I assume that experiences of being free to do otherwise have phenomenal content that is inaccurate if determinism is true, just as libertarians claim. Then I argue that such experiences also have phenomenal content that is accurate if determinism is true. So an experience with libertarian content can be accurate under determinism. Second, I argue that implicitly libertarian beliefs can be accurate, even assuming determinism. Only when one makes an explicit

incompatibility judgment is one's belief false. Thus, implicitly libertarian beliefs are not incompatibilist. My view does not provide a full compatibilist theory of the ability to do otherwise. Still, on my view our experiences and beliefs concerning such freedom are consistent with determinism. My view also explains any temptation we may feel to judge our experience and beliefs as inconsistent with determinism.

Preface

Chapter 2 is based on a co-authored paper with Matt Bedke and Shaun Nichols, with Oisín Deery as first author. The identification of the research program and the primary design of the experiments conducted as part of the research were undertaken by Oisín Deery. The training-to-criterion for determinism that we used in the experiments was introduced by Shaun Nichols, and was subsequently developed by all three authors. Data analysis on the experiments conducted in Study 1 was performed by Matt Bedke. All subsequent data analysis on studies 2 and 3 was conducted by Oisín Deery. Preparation of the manuscript was performed by all three authors.

At the time of submission of this dissertation, Chapter 2 has been published in D. Shoemaker (ed.), *Oxford Studies in Agency and Responsibility: Vol. I*. This material is reprinted with the permission of Oxford University Press. It has been altered in various ways for the sake of maintaining consistency with the rest of the dissertation.

For the experiments conducted as part of the research reported in Chapter 2, approval was obtained from the Behavioural Research Ethics Board at the University of British Columbia. The Certificate Number of the Ethics Certificate obtained is H10-01096.

Table of Contents

Abstract	ii
Preface	iv
Table of Contents	v
Acknowledgments	viii
1 Introduction	1
1.1 How to Think about Free Will	1
1.2 Determinism	6
1.3 Free Will	13
1.4 The Experience of Freedom	26
1.5 Outline of the Project	31
2 Phenomenal Abilities: Incompatibilism and the Experience of Agency 32	
2.1 Background	33
2.1.1 Introduction	33
2.1.2 The Experience of the Ability to Do Otherwise	34
2.1.3 Previous Work on the Phenomenology of Free Will	37
2.2 Study 1	40
2.3 Study 2	46
2.4 Study 3	51

2.5	General Discussion	56
2.5.1	Incompatibilism	56
2.5.2	The Ability to Do Otherwise	57
2.5.3	Misinterpreting One’s Agentive Experience	59
2.5.4	Deliberation Compatibilism	61
2.6	Conclusion	63
3	Is Agentive Experience Compatible with Determinism?	65
3.1	Introduction	65
3.2	Agentive Experience	67
3.3	Horgan’s View	72
3.4	Problems with Horgan’s View	78
3.5	Problems with Horgan’s Contextualism	82
3.6	An Alternative Compatibilist Proposal	86
3.7	Conclusion	90
4	Against an Argument for Libertarianism	92
4.1	Introduction	92
4.2	Compatibilist Strategies	93
4.3	An Analogy with Color	95
4.4	Libertarian Content	99
4.5	Compatibilist Content	101
4.6	The Two-Stage View	105
4.7	Conclusion	107
5	Indeterminism, Experience, and Compatibilism	109
5.1	Introduction	109
5.2	Do Indeterminist Beliefs Have Their Source in Experience?	112
5.3	A Forward-Looking Model	119
5.4	Etiology of Indeterminist Experience	123
5.4.1	Prospection is Forward-Looking	123

5.4.2	Prospection Can Be Experienced	126
5.4.3	Prospected Choices Are Free Variables	127
5.4.4	Indeterminist Experience and Judgment	132
5.5	Belief in Indeterminist Freedom is Not Justified	134
5.5.1	Radical Compatibilism	136
5.5.2	Contextualism Revisited	138
5.6	Conclusion	144
6	Conclusion	146
6.1	What the Dissertation Achieves	146
6.2	New Evidence on the Descriptive Question	149
6.3	Against a Compatibilist Error Theory	149
6.4	An Alternative Compatibilist Proposal	151
6.5	The Source of Libertarian Content	154
6.6	Conclusion	157
	Bibliography	158
A	Survey Materials for Chapter 2	172

Acknowledgments

To my wife, Katherine Bailey: thank you for your encouragement and continued belief in me. I do not know whether I can ever repay my debt of gratitude to you; but I will try.

I thank my advisor, Paul Russell, and my dissertation committee, including Derk Pereboom (Cornell University) and Murat Aydede for their advice and guidance. I thank Matt Bedke, also on my committee, and, at the University of Arizona, Shaun Nichols, for their critical input to Chapter 2. At the University of Arizona, I thank Terry Horgan, Michael McKenna, and Uriah Kriegel for their support during my stay there, and Jenann Ismael for her inspiring ideas. Thanks to Mark Balaguer at California State University, Los Angeles, for his suggestions regarding Chapter 4, and to Peter Railton at the University of Michigan, Ann Arbor, for correspondence that proved crucial to Chapter 5. For advice on how to write philosophy, my thanks to Dominic McIver Lopes at UBC. Finally, I thank my examination committee, especially Jonathan Jenkins Ichikawa and Eddy Nahmias, who (in addition to those listed above) gave me valuable comments for my final revisions.

I was supported at UBC by a Graduate Entrance Scholarship, a UBC Graduate Fellowship, a Tina and Morris Wagner Foundation Fellowship, and (for one year) by a Four-Year Doctoral Fellowship. I was also supported by a Pacific Century Graduate Scholarship, a joint award administered by UBC and The Ministry of Advanced Education, Province of British Columbia. In 2010, I was supported by a Doctoral Fellowship awarded by the Social Sciences and Humanities Research Council of Canada. In 2011, I held a Visiting Graduate Fellowship at Cornell University.

Chapter 1

Introduction

I've visited thirty-one inhabited planets in the universe. Only on Earth is there any talk of free will.

— Kurt Vonnegut, *Slaughterhouse-Five*

1.1 How to Think about Free Will

Take ‘free will’ to mean the freedom to choose among alternative possibilities,¹ and ‘determinism’ to mean that, given the past and the laws of nature, there is, at any given time, just one physically possible future.² Indeterminism is the opposite of this thesis: holding fixed the past and laws, there is more than one physically possible future.

Do we have free will? To answer this question, we must clarify what we mean by ‘free will,’ even in the sense of being free to choose among alternatives. Is freedom compatible with determinism? Or with indeterminism? Let us focus on the first question.

There are several reasons why one might be interested in free will. First, one might be interested in how to distinguish actions done from weakness of will

¹ I discuss other senses of ‘free will’ later.

² I also discuss the thesis of determinism in greater detail later.

from actions performed under compulsion. Here, the question whether freedom is compatible with determinism may not even arise. Alternatively, one might be interested in the sort of freedom-relevant control required for moral responsibility, quite apart from whether such control demands that agents be able to choose among alternatives.³ Finally, one might take it as a basic insight that being free to choose among alternatives for action, and thus being able to do otherwise, is fundamental to the notion of free will. Moreover, one might think that our possessing such an ability is confirmed by our experience of choosing. In this dissertation, I am concerned exclusively with the latter reason for being interested in free will, as I explain below.

There are a number of standard positions that one can occupy about free will. Incompatibilists think that free will is incompatible with determinism. For incompatibilists, the fact that determinism entails that there is only one physically possible future means that we lack alternatives for action. So, if free will requires being able to choose among alternatives, free will is inconsistent with determinism. Some incompatibilists—the libertarians—maintain that we have free will, and as a consequence they think that determinism is false. The task for libertarians, as we shall see, is to show how indeterminism is supposed to help with free will.⁴ Other incompatibilists—traditionally called hard determinists—think that we lack free will *because* determinism is true. Today, though, few philosophers are willing to commit to the truth of determinism. Instead, the current inheritor view of hard determinism is hard incompatibilism, whose advocates maintain that free will is incompatible with determinism, but is also probably incompatible with indeterminism. As a result, most likely we lack free will whether determinism is true or false.⁵

³ Many philosophers now think that free will—at least when thought of as the sort of freedom or control required for moral responsibility—does not require having alternative possibilities. I discuss this issue later.

⁴ Since I focus on compatibility with determinism, I will not discuss further the question whether free will is compatible with indeterminism, other than by way of a few brief comments in the next paragraph.

⁵ There is also the “impossibilist” view, according to which free will (in the sense required for responsibility) is impossible, since it is ruled out by both determinism and indeterminism (e.g., G. Strawson 1994).

By contrast, compatibilists think that free will is consistent with determinism. As a result, if free will requires being able to choose among alternatives, then compatibilists must show how choosing among alternatives is possible, assuming determinism. Typically, compatibilists also argue that libertarian views are incoherent, or deviate from a naturalistic account of the world, or fail to show how indeterminism provides for more control than determinism does. Compatibilists who press this latter point may further argue that indeterminism *hinders* control. After all, one might think that for agents to control their decisions, their decisions must flow smoothly from their beliefs and desires. Indeterminism threatens this picture. For, when we hold fixed the entire history of an indeterministic world (including the internal states of the agent) right up until the moment of the agent's decision, what the agent ends up deciding is indeterminate right up until the moment of her decision. This is so *despite* the agent's beliefs, desires, and reasons for deciding one way rather than another. Such "randomness" threatens to undermine agents' control over their decisions, and thus free will—or so compatibilists claim. Traditionally, compatibilists have also thought that determinism is *true*, since they thought that undetermined events are uncaused, which seemed absurd. Given that events *are* caused, compatibilists thought, events are determined. Today, however, most compatibilists accept that indeterminism might be true. Despite this, they insist that agents have free will only if indeterminism plays no control-undermining role in the production of decisions. Moreover, even if determinism is true, that is no threat to freedom.

I focus primarily on the question whether free will is compatible with determinism. This question—the compatibility question—has come to dominate the literature on free will in recent decades. Many philosophers think that the way to answer this question is to show whether there is a possible world in which determinism is true, yet in which an agent is able to do otherwise. If so, compatibilism is vindicated; otherwise not. That is not primarily how I will approach matters. At least at first, I will focus on the "natural" compatibility question, which is a de-

scriptive question about people’s pre-theoretic belief-tendencies about their own and others’ agency. The natural compatibility question asks whether people *begin* as compatibilists, prior to their consideration of philosophical theories. Of course, even establishing whether people are natural compatibilists in some respect (for instance, about the freedom to do otherwise) would leave open the question how we *ought* to think about our agency in that respect—about whether we *should* be theoretical compatibilists, say. Even so, it may inform the approach that we take when addressing that question.

Whether people are natural compatibilists in some respect is straightforwardly an empirical question, which I will call the descriptive question. Once we have answered this question, we then address the substantive question: are people *in fact* agents of the sort that they tend to believe they are (in a given respect)? Finally, there is the prescriptive question, which asks how we should best theorize about free will (in a particular respect), given how we have answered the descriptive or substantive questions. If the answer to the descriptive question is that people tend to believe that they have a sort of freedom that requires conditions G, H, and K to be satisfied, yet it is reasonable on independent grounds to think that only K is actually satisfied, then our best prescriptive theory of freedom may have to depart from our pre-theoretic beliefs.⁶

There are a number of underlying issues here—related, for example, to philosophical methodology, the reference of concepts, and so on—which I want to bracket. In what follows, I mostly adopt the approach of methodological naturalism.

As I see things, methodological naturalists (henceforth, “naturalists”) are pri-

⁶ Others who think about the compatibility question in roughly this way are Nichols (2008), Vargas (2013), and Balaguer (e.g., 2004). On Balaguer’s way of viewing matters, the *what-is-free-will question* is a semantic or conceptual question about the meaning of ‘free will.’ Unlike Balaguer, I avoid putting things in semantic terms, since ‘free will’ (or ‘freedom’) most likely does not have a single meaning. As with any term that has been in our language for some time, there are many (perhaps overlapping) senses of the term, and it seems unlikely that any one of these is *the* meaning of that term.

marily concerned with explaining aspects of the world. To this end, they begin with a target phenomenon. In order to establish what the phenomenon's characteristic features are, naturalists rely on a combination of (i) ordinary judgments or belief-tendencies about supposed cases of the phenomenon, (ii) other relevant background judgments or beliefs, and (iii) supplementary empirical evidence. Naturalists think that we should establish what the contents are of the judgments mentioned in both (i) and (ii) empirically, by finding out what people's belief-tendencies actually happen to be. Empirical methods also bear on this procedure in another way, by enabling us to determine whether agents' judgments have been produced by reliable mechanisms. If not, such judgments may be explained away by an error theory. Also taken into considerations from the outset is (iii) supplementary evidence not directly related to establishing what our beliefs are or to assessing their reliability, yet which is otherwise relevant to the target phenomenon (see e.g., Paul 2010 for an overview of this method).

We then build an explanatory theory of the phenomenon. In doing so, we treat all our judgments, as well as our explanatory hypotheses, as defeasible in light of new findings. As we proceed, each part of our theorizing is constrained at each step by empirical evidence. In this way, we arrive at theories of the target phenomenon, and we decide among them (if necessary) by inference to the best explanation.

In the chapters that follow, I conclude that people's experience of freedom is, in a sense, incompatibilist. People tend to experience possessing (and thus to believe that they possess) a sort of freedom to do otherwise that they cannot possess if determinism is true. As a result, if determinism is true there seems to be a gap between the sort of freedom that people experience possessing and the sort of freedom that they do possess. Despite this, I defend a compatibilist answer to the prescriptive question about experiences of freedom, by showing how such experience is, in an important sense, usually veridical (and therefore compatibilist), even assuming determinism. I also propose an explanation for how incompatibilist experiences are generated, and in such a way that we are

not warranted in believing libertarianism simply on the basis of experience, as some libertarians have suggested that we are. I also maintain that our freedom to do otherwise is compatibilist. In this way, I defend natural incompatibilism about the descriptive question, but normative compatibilism about the prescriptive question.⁷

In the next two sections, I say more about the items whose compatibility is at issue in debates about free will: determinism and freedom. After that, I say something further about each of the descriptive and prescriptive questions, as well as outlining why I focus on the experience of freedom. In the last section, I provide an overview of the thematic and argumentative structure of the subsequent chapters.

1.2 Determinism

The thesis of determinism that concerns us is nomological determinism. This is the thesis that all events are determined to occur by natural—usually physical—laws. As Kadri Vihvelin notes (2011; I rely heavily on Vihvelin’s article in what follows), this is a contingent thesis about the laws of nature. It says, first, that the laws are not probabilistic. Deterministic laws entail exceptionless regularities, such as that all Fs are Gs—i.e., Fs have an objective probability of 1 of being Gs.⁸ By contrast, probabilistic laws state that Fs have an objective probability of less than 1 of being Gs. Second, determinism says that the totality of the laws is all-encompassing—the totality of the laws applies to everything in the universe, not just to a part of it. So, determinism is defined as follows: the conjunction of a complete statement of the non-relational facts of the world at a time, t , with a complete statement of the laws of nature, entails all other non-relational truths about the world at times other than t , including truths about future human deci-

⁷ I do not directly address the substantive question; I leave it aside as work for another day.

⁸ Although determinism, as defined here, does not require a non-Humean account of laws, I will assume that the laws are non-Humean, as I explain in the next paragraph. I reassess this assumption in Chapter 5.

sions and actions.⁹ The main consequence of this thesis is that at any given time there is only one physically possible future, given the past and the laws—there is just one way that it is physically possible for the world to unfold.¹⁰

Even when it comes to all-encompassing deterministic laws, there is a further important distinction to be made between Humean and non-Humean laws.¹¹ The difference lies in whether the truth of a law is established by events that fail to disconfirm it, or whether something else establishes the truth of a law, so that events conform to it. The first account of laws is Humean, while the second is non-Humean. This is important, since on a Humean account if determinism is true then a law that determines whether an agent, S, performs an action, A, at a particular time, *t*, does not “settle,”¹² as John Perry puts it (2004), whether S A-s at *t*, since S’s A-ing at *t* is part of the sequence of events that establishes the truth of a law, not something that is settled by it. According to Perry, on a Humean account of laws, “What we do is up to us; laws are merely descriptions of what we do that will end up being true once human activities are complete. Laws determine, but do not settle” (2004: 239). By contrast, on a non-Humean account, if determinism

⁹ Vihvelin (2011) identifies three ways a world might be indeterministic or non-deterministic: (i) some of its laws may be probabilistic, (ii) the totality of its laws may not be all-encompassing, or (iii) it may have no laws. Vihvelin only calls worlds that are non-deterministic in the first way “probabilistic worlds,” since these worlds have probabilistic laws. Worlds that are non-deterministic in the second way she calls “partly lawless,” while those that are non-deterministic in the third way are “lawless.”

¹⁰ I will say more in Chapter 5 about whether this is a good way to understand determinism. It is, at any rate, the standard way.

¹¹ For broadly Humean, non-necessitarian accounts of the laws of nature, see Lewis (1973; 1983; 1994) and Earman (1984). For non-Humean, necessitarian accounts, see Armstrong (1978; 1983), Dretske (1977), and Tooley (1977).

¹² Here is how Perry explains what he means by ‘settled.’ He outlines three sorts of propositions: (1) those that are made true by events (e.g., in the way that the proposition ‘I was born in Dublin’ is made true by the event of my actually having been born in Dublin), (2) those that are not made true by events (e.g., Pythagoras’s theorem), and (3) those that while not yet *made true* by events nonetheless have their truth value settled. If the propositions describing the laws of nature are not made true by events (i.e., if the laws are like Pythagoras’s theorem, and thus are non-Humean), then if descriptions of the laws of nature and the facts of the past *entail* another proposition, *p*, about the future, the truth value of *p* is *settled*, in Perry’s parlance, even though *p* has not yet been made true by events.

is true then the laws together with the facts of the past not only entail facts about the future, but also settle them.

Many philosophers, Perry included, find Humean accounts of laws unconvincing. As Perry puts it, “It seems to me much more plausible that . . . [a] . . . law gets at something (or some things) about the universe that explains why things conform to the law and it has no disconfirming instances” (2004: 240). I will assume a non-Humean viewpoint, since it prevails in the literature on free will. Future events are settled, in Perry’s parlance, by the laws; events conform to the laws, rather than the laws’ being made true by the sequence of events.¹³

It is also important to distinguish determinism from fatalism (see e.g., Bernstein 2002). Unlike determinism, fatalism does not count our deliberations as necessarily exerting any causal influence on events. Fatalism is the thesis that whatever is fated to happen will happen no matter what one does, or—as Peter van Inwagen puts it—“*no matter what choices and decisions one makes*” (1983: 25, n. 3).¹⁴ Determinism, by contrast, implies that one’s deliberations and actions do exert an influence on events.

Along with fatalism, there are several other theses with which determinism might be confused. It is useful, therefore, to note what else determinism does *not* say. First, determinism does not say anything about the predictability of events. In particular, it does not say that, in virtue of their being deterministic, events are easy to predict. Chaos or complexity theory says that events in complex deterministic systems are difficult to predict. Yet, according to some interpretations of quantum mechanics, certain behaviors of indeterministic systems are easy to predict (see e.g., Earman 2004).

Second, determinism is not a thesis about causation. Here, I mean two things: (a) causation need not be deterministic, and (b) determinism need not—and arguably *should* not—be stated in terms of causation at all. Both points are important, since especially in the older literature on free will (e.g., Hobart 1934;

¹³ Again, I revisit this issue in Chapter 5.

¹⁴ Even indeterministic fatalism seems to be coherent, at least when the thesis applies only to some events.

Ayer 1954) it was often assumed that determinism just *is* the thesis that every event has a cause. By contrast, it is now widely accepted that the laws may be probabilistic, and causation may be probabilistic too: events might have causes even if determinism is false.¹⁵ The conceptual possibility of indeterministic causation was noted decades ago by Elizabeth Anscombe (1981/1971). According to Anscombe, we often claim to know whether one event caused another, even when we do not know whether the second event was determined by the first event. Moreover, the concept of causation seems to be of “giving rise to,” whereas the concept of determinism is of “having to give rise to.” So indeterministic causation seems conceptually coherent.¹⁶

I maintain that we can (and arguably should) think of determinism in non-causal terms in any case—the prevalence in the free-will literature of talking about ‘causal determinism’ notwithstanding.¹⁷ John Earman argues that we should drop the term ‘causal’ from discussions of determinism, since to speak of ‘causal de-

¹⁵ Two points are worth noting here. First, Vihvelin observes that, “[I]t might be true that every event has a cause even if our world is neither deterministic nor probabilistic. If there can be causes without laws (if a particular event, object, or person can be a cause, for instance, without instantiating a law), then it might be true, even at a lawless or partly lawless world, that every event has a cause” (2011). Second, Vihvelin also points out that it is unclear “whether determinism entails the thesis that every event has a cause. Whether it does so or not depends on what the correct theory of causation is; in particular, it depends on what the correct theory says about the relation between causation and law” (2011).

¹⁶ Of course, the possibility that *indeterminism* might be true goes back even further, to the development of quantum mechanics in the 1920s. According to quantum indeterminacy, a physical system does not have any determinate state that determines the values of its measurable properties. The system has only a defined set of probable states. On the Copenhagen interpretation of quantum mechanics, this probability is objective. The process of measuring the properties of the system results (somehow or other) in just one of these objectively probable states becoming instantiated. In effect, indeterminacy is built right into the fabric of the universe. On this interpretation, there can be no laws in the sense of exceptionless regularities stating that all Fs are Gs (i.e., with a probability of 1). (On other interpretations of quantum mechanics, such as the “hidden-variables” or the “many-worlds” interpretations, the apparent indeterminacy in quantum mechanics is merely apparent and can be accounted for deterministically.) Yet, while an indeterministic theory of causation may rely on, say, the truth of the Copenhagen interpretation of quantum mechanics, quantum mechanics stands in no need of a theory of causation. In this way, as I explain below, whenever we speak of determinism or indeterminism in quantum mechanics, we are speaking in terms of an *entailment* notion of determinism, not a causal one—where entailment is understood broadly enough to include mathematical consequence.

¹⁷ A caveat to this claim will be introduced in Chapter 5.

terminism' is just "to explain a vague concept—determinism—in terms of a truly obscure one—causation" (Earman 1986: 5).¹⁸ If instead we think about determinism in entailment terms, then determinism requires only that (i) there be a well-defined state or description of the world at a time, t , and (ii) there be exceptionless laws of nature that are true at all times and places. If (i) and (ii) logically entail complete descriptions of the world at times other than t , then determinism is true; otherwise not. Although (i) is false if the standard interpretation of quantum mechanics is correct (see footnote 16), and (ii) is false if the laws of nature are probabilistic (or if there are no laws), all that is required to articulate the thesis of determinism other than (i) and (ii) is the notion of entailment, which should be understood broadly enough to include mathematical consequence as it appears in theoretical physics.¹⁹

Third, and finally, determinism neither entails nor is entailed by physicalism, the thesis that everything is physical or is (in some sense) necessitated by the

¹⁸ See also van Inwagen (1983: 65).

¹⁹ A slight wrinkle here is that, as Derk Pereboom notes (unpublished handout), Frankfurt cases appear to show that preceding conditions might entail that an agent, S, performs an action, A, without playing any part in actually bringing it about that S A-s: the preceding conditions entail that S A-s, yet they do not cause A (and so they do not causally determine A). This is supposed to be a problem for the entailment view of determinism since we are meant not to assume determinism in Frankfurt cases, yet the entailment view appears to do just that, if S's A-ing is entailed by the preceding conditions. Thus, we should prefer a causal formulation of determinism. In response, recall my definition of determinism: the conjunction of a complete statement of the non-relational facts of the world at a time, t , with a complete statement of the laws of nature, entails all other non-relational facts about the world. However, the preceding conditions to which Pereboom alludes are, it seems to me, relational facts. So, by adopting an entailment view of determinism, it is not true that one begs the question against incompatibilists in Frankfurt cases. To explain: Consider that it is a hard (non-relational) fact relative to 2p.m. that Jones eats noodles at 1p.m. By contrast, it is a soft (relational) fact relative to 2p.m. that Jones eats noodles for two hours after having started to eat them at 1p.m. Now, the relevant fact in a Frankfurt case at t_1 is meant to be something like this: If Black sees a prior sign at t_2 indicating that Jones is about to choose to B at t_3 , then Black will intervene to ensure that Jones A-s at t_3 . That seems like a relational fact. Thus, it is not the kind of fact that features in a statement of determinism. So, even if such a fact entails that Jones A-s at t_3 , it does not thereby illicitly assume determinism. More importantly, the facts appealed to in the thesis of determinism are stated in the language of physics: they are physical facts. The fact that supposedly entails that Jones A-s at t_3 is not a physical fact. Thus, even if such a fact entails that Jones A-s at t_3 , it does not thereby illicitly assume determinism.

physical. Peter van Inwagen (2000) thinks that the problem of free will is so abstract that it must arise in roughly the same form even in a world inhabited only by immaterial beings: either such beings obey deterministic laws, or they do not.²⁰ Conversely, there are also possible worlds in which determinism is false, yet physicalism true (our own may be such a one). However, the form of nomological determinism that is generally thought to threaten human freedom is stated in terms of physical laws.

Henceforth, whenever I speak of determinism I will have in mind nomological determinism, which I take to be a thesis best stated in terms of logical entailment and physical laws.²¹ At bottom, this thesis says—minimally—that there are no objectively probabilistic physical laws governing our world, and, as a result, objectively probabilistic events do not occur.²² The main consequence of this thesis (if it is true) is that there is only one possible way for the actual world to unfold, physically speaking.

Why worry about determinism? After all, the standard interpretation of quantum mechanics is indeterministic.²³ On this view, our world is indeterministic at least in the sense that some of its physical laws are objectively probabilistic. If so, determinism is false and is no threat to freedom. One response to this proposal is to say that even if quantum indeterminism is true, human behavior may still be “near-determined”:²⁴ it may be that quantum indeterminacies make no difference at the macro level of humans or neurons (or whatever else). In this way, physics might be indeterministic, while neuroscience, for instance, is nevertheless near-deterministic.

²⁰ Clearly, however, these laws will not—however they are to be stated—be stated in terms of physical theories.

²¹ Again, an amendment of this way of thinking about determinism will be introduced in Chapter 5.

²² A minor caveat: some non-Humean, necessitarian accounts of laws are driven by a theory of dispositions, “according to which dispositions have their causal powers essentially. . . . Laws, then, are entailed by the essences of dispositions” (Carroll 2011; Cf. Bird 2005). On this view, causation is a more fundamental notion than law. Thus, causation may enter into an account of nomological determinism via the back door, so to speak—that is, via a dispositional account of natural laws. In what follows, I leave aside this technicality.

²³ See footnote 16 above.

²⁴ The term ‘near-determined’ is introduced in this context by Honderich (1993).

There is an important distinction to be made here. First, it might be that quantum-level indeterminacies do not have any effect at the neural level, so that determinism is false when stated in terms of physical laws, yet true when stated in terms of neuroscientific laws. In that case, neural determinism is true. So there is no need to speak of near-determinism.²⁵ Second, it might be that quantum-level indeterminacies do, as a matter of fact, have effects at a neuronal level, but not anything like as often—or in such a way—as to allow for indeterminism in decision-making.²⁶ In that case, neural near-determinism is true. Even so, it appears that we have nearly as much reason to worry about near-determinism as we do to worry about determinism, since neural near-determinism presents almost as much of a *prima facie* threat to free will as does determinism, yet is much more likely to be true.

Alternatively, the world might be near-deterministic in the sense of being imperfectly deterministic. That is, the universe (or our region of it) might be deterministic for long periods of time (say, millions of years), yet occasionally there might occur spontaneous particle-creation events that violate physical determinism (Cf. Hofer 2010). Thus, determinism could be false as a general thesis about our universe. Nevertheless, our world would be sufficiently deterministic to pose a *prima facie* threat to free will.

²⁵ To see how quantum indeterminism might effect decision-making, see Dennett (1978: 46–47). To paraphrase: Imagine that my correctly answering No to an easy Yes/No question either by pushing a pedal with my left foot or by pushing a button with my left finger (where either action would register the correct answer) depends on the output of a quantum-indeterministic randomizer. Say I want to answer No. Now the randomizer outputs an instruction telling me whether I should press the pedal or push the button. This output—and therefore my action of, say, pressing the button—is (per hypothesis) undetermined, even if intrinsically I am a deterministic system. The question is whether such amplification of indeterminism from the micro- to the macro-level is possible without such a device.

²⁶ Mere indeterminism, even if it occurs at the moment of choice, seems insufficient for free will, even for incompatibilists.

1.3 Free Will

According to one way of thinking about human agency, free will is *moral freedom*, or the “freedom-relevant condition necessary for moral responsibility” (Campbell et al., 2004: 2). We possess such freedom when we control our actions in the strongest manner necessary for being morally responsible for them.²⁷ In other words, to judge an agent, S, as morally responsible for performing an action, A, it must have been the case that S, when she A-ed, exercised this necessary form of control over her A-ing.

In thinking about the sort of control required for moral freedom, it is standard to distinguish between leeway and sourcehood views. Leeway views require that an agent be able to do otherwise in order to be morally responsible, while sourcehood views do not. Sourcehood theorists argue, instead, that an agent must in some relevant sense be the source of her actions in order to be responsible for them, where this requirement is spelled out in a way that does not require that the agent have been able to do otherwise. On another way of thinking about matters, free will is (or requires) being able to do otherwise. Call such freedom *modal freedom*.²⁸ It is a further question, on this taxonomy, whether moral freedom requires modal freedom. Traditionally, it was held by almost all parties to debates about free will that modal freedom *is* a necessary condition on moral freedom. The existence of sourcehood views demonstrates that this is no longer the case.

Notice, too, that we can also think about free will independently of responsibility, solely in terms of human freedom and abilities. By analogy, consider our interest in personal identity. One reason that the topic of personal identity is important to us is because of its relevance for moral responsibility. We want to know what the criteria are for an agent’s remaining the same person over time because we want to be able to connect the person whom we apprehend today for

²⁷ Thus, there could be conditions sufficient for the control required for moral responsibility that are stronger than this.

²⁸ Here, I follow the example of Holton (2010).

the crime of murder to the same person (according to the criteria) who committed that murder yesterday. However, this is hardly the only reason we are interested in personal identity. Another reason is that we want to know whether we are the kind of creature that we conceive ourselves to be. This question presses on us in its own right. It may turn out that there are no viable criteria for personal identity. Such an answer would impact what we believe about ourselves, and may undermine a view of ourselves that we value, independently of any further considerations regarding moral responsibility.²⁹ Likewise, I suggest, modal freedom is important to us independently of moral freedom. Perhaps we tend to think that we possess a sort of freedom or ability that, in fact, we lack. If so, this may impact what we believe about ourselves, and it may undermine a view of ourselves that we value, perhaps deeply, and this may be so quite independently of any further value-relevant considerations to do with responsibility.

If we *do* think about free will independently of responsibility, just in terms of freedom and abilities, then we are interested solely in modal freedom, which is arguably more fundamental than moral freedom anyhow, since, as Campbell and colleagues note (2004), the difficulties regarding modal freedom can be stated without reference to moral responsibility.³⁰ In what follows, I am concerned only with modal freedom.

A traditional analysis of free will, conceived as modal freedom, begins as follows:

- (1) S has free will only if S is able to do otherwise than she does.

Here, modal freedom is claimed to be (at least) a necessary condition on free will. Of course, if free will *just is* modal freedom (as e.g., van Inwagen 2008 maintains),

²⁹ See Parfit (1971; 1987) and Velleman (2006) for such views about personal identity.

³⁰ One difficulty regarding modal freedom is characterizing what it *is* to be able to do otherwise. I should also note that it is not obvious whether modal freedom *can* be characterized without reference to moral notions. Van Inwagen (2000: 17–18) characterizes what it is for an agent, S, to be able to A partly in terms of whether S is in a position to promise that she will A. Yet, promising is a moral notion.

then the ability to do otherwise is both a necessary and a sufficient condition on freedom. Thus:

(2) S has free will iff S is able to do otherwise than she does.

In either case, the freedom at issue is the ability to decide between mutually incompatible courses of action. If we think about free will independently of moral responsibility—as what I have been calling modal freedom—then whatever we think about the requirements of moral responsibility, we will still want to know whether we are free to do other than we do (and thus whether we have free will). As I mentioned earlier, in order to answer this question we must first establish whether being free to do otherwise is compatible with determinism. One influential compatibilist thought is that freedom should be understood in contrast to constraint, or coercion. According to this proposal, an agent is able to do otherwise just in case, if she had chosen, or wanted, or tried to do otherwise, then she would have done so (Cf. Moore 1912). By contrast, incompatibilists think that being able to do otherwise requires being free to do something other than what one does, all prior conditions (including one’s beliefs, desires, etc.) remaining the same.

We are now in a position to see why ‘free will’ is actually an unhelpful term. When we ask whether an agent is free in a given situation, our answer depends on what we mean by ‘free.’ Even in debates about free will, it may be unclear whether we mean moral freedom or modal freedom (or: whether the latter is a condition on the former). I suggest that instead of talking about free will, we should instead talk about the various more fine-grained distinctions and notions that are of interest in such debates. When asking compatibility questions, I think we do better to ask about the consistency of determinism and *moral responsibility*, or *sourcehood control*, or *the freedom to do otherwise*, and so on. Once we distinguish these questions, we may answer some of them differently than others. To the question whether the conditions of moral responsibility are compatible with determinism,

our answer might be “Yes,” while our answer to the question whether the freedom to do otherwise is compatible with determinism might be “No.” Thus, we could be compatibilists about responsibility, yet incompatibilists about the freedom to do otherwise (or vice versa).³¹ Whether we are *natural* compatibilists about any of these notions and determinism will, once we hold fixed the definition of determinism, depend entirely on our presuppositions (or belief-tendencies) about the other phenomenon.³²

An account of the ability to do otherwise is required for any adequate account of ourselves as agents. Even sourcehood theorists, who deny that the freedom to choose among alternatives is required for moral responsibility, still presumably think that sometimes we are able (in some sense) to do otherwise.³³ We need an account of such freedom. One of the first questions that arises in developing such an account is:

The freedom-to-do-otherwise-compatibility question: Is the freedom to do otherwise compatible with determinism?

One powerful reason for thinking that the freedom or ability to do otherwise is *not* compatible with determinism is the consequence argument. This argument relies on the notion of an agent’s lacking power over a fact. For an agent, S, to lack

³¹ See e.g., Fischer (1994, 2007, 2013) and Fischer and Ravizza (1998) for a version of the former sort of view in the philosophical literature.

³² When it comes to the intuitions people have about free will (or moral responsibility, or the ability to do otherwise, or determinism, etc.), it is worth noting that, psychologically-speaking, one should be hesitant to conclude categorically that any individual person either does or does not have incompatibilist intuitions. According to empirical results obtained by Deery and colleagues (in preparation), people’s pre-theoretic folk intuitions about free will are often naturally both incompatibilist and compatibilist. Deery and colleagues found that when free will is understood as an agent’s being the ultimate source of her actions, for instance, participants in their studies agreed with incompatibilist statements. Yet, participants also agreed with compatibilist statements of this conception of free will, and to a similar extent. Thus, all else being equal, when respondents were not placed in experimental situations that required them to resolve an explicit conflict between opposing intuitions, they possessed both incompatibilist and compatibilist intuitions about being the ultimate source of their actions.

³³ A deeper issue is whether incompatibilist sourcehood theorists require alternatives covertly, by requiring indeterminism.

power over a fact, A, is for S to be unable to act in such a way as to ensure the falsity of A. Furthermore, if S lacks power over A, and lacks power over whether another fact, B, follows from A, then S lacks power over B. Powerlessness transfers from a fact to its consequences. This principle is central to the consequence argument, as the argument's name suggests. Recall that determinism implies that S's action at *t* is a consequence of the distant past and the laws.³⁴ If S lacks power over the distant past and the laws, and S lacks power over whether her action at *t* follows from the past and the laws, then S lacks power over her present and future actions.

There is a vast literature on this argument, which I want to bracket. According to the way in which I suggest that we approach the freedom-to-do-otherwise-compatibility question, any answer to this question will depend on what people actually tend to believe about their abilities. In other words, we must first answer the descriptive question about people's relevant belief-tendencies. Holding fixed the definition of determinism, the answer to the freedom-to-do-otherwise-compatibility question—understood as the *natural* compatibility question—depends solely on the answer to this descriptive question. In this way, we find out whether people are natural compatibilists about freedom. So the answer to the freedom-to-do-otherwise-compatibility question depends on the following question:

The freedom-to-do-otherwise-descriptive question: What sort of freedom do people tend to believe they have?

How should we answer this question? One way is simply to ask people about their belief-tendencies. Another way, which I will adopt, is to ask people about their *experience* of being free to do otherwise. Appeals to the experience of freedom have a long history in free-will debates, going back at least to Thomas Reid (1788), who argued that our experience of incompatibilist freedom justifies belief

³⁴ More carefully, determinism implies that a statement describing an action of S's is a consequence of a conjunction of a statement of the laws and a statement of the facts of the world at a time in the distant past.

in libertarianism. Compatibilists, too, have appealed to our experience of freedom in order to deny such incompatibilist claims: they insist instead that our experience carries with it only *compatibilist* commitments. In Chapter 2, I present evidence showing that, as it turns out, people describe their experience of freedom as incompatibilist. Before turning to these issues, however, it will be helpful to say a few words about the sort of freedom that is at stake.

First, we need to distinguish alternative possibilities from the freedom to do otherwise. Having the latter entails having the former, but not vice versa. Even if indeterminism is true and we have indeterministic alternatives, we might lack the ability to do otherwise. Imagine that you are in a runaway car, in which the steering mechanism and brakes are not working. Here, there might exist alternatives—the car might go, even indeterministically, in various directions—but you lack the ability to make it go in any direction at all, or even stop. This point also applies if we assume determinism, at least on the assumption that there are compatibilist alternative possibilities. (I will call both compatibilist and incompatibilist alternatives *metaphysical*, in order to avoid begging the question against compatibilists by assuming—as is done sometimes—that only incompatibilist alternatives are metaphysical.)

We must also distinguish metaphysical from epistemic possibilities. Unlike the former, the latter are uncontroversially compatibilist. For an event to be epistemically possible is just for an agent not to know whether that event will occur. Consider again the example of the runaway car. Given that you do not know whether the car will go left or right, its going in either direction is epistemically possible, even if one direction is disallowed by the (deterministic or indeterministic) laws of nature.

Regarding ‘can,’ I will use ‘can do otherwise’ as predicative of an ability (which may or may not be compatibilist) to choose among metaphysical, not merely epistemic, alternative possibilities. This usage differs from that which

indicates moral or legal *permissibility*.³⁵ Using ‘can’ in a permissibility sense, you might insist that “You can’t do that!” where (for instance) I am in the process of stealing a car. What you mean is that my stealing the car is morally or legally impermissible.³⁶

There are several other common senses of ‘can’ that differ from the one in which I am interested. One of these is predicative of a *disposition*, another of a *power*, and yet another of a *capacity* or *general ability*. I will take each of these in turn.

When we say that a cup can break, what we mean is that it is disposed to break when struck. This usage of ‘can’ attributes a disposition. On some views, a power is not essentially different from a disposition: a power is simply a disposition to produce a specific manifestation under certain circumstances. Thus, dispositions are causal properties, which are identical to their causal bases, e.g., *being fragile* just is *having a certain physical structure*. In this sense, a wineglass has a power, in being fragile. Some dispositions, however, clearly count as powers in a different sense. Consider a steel hammer, which has a power in the aforementioned sense by having a high tensile strength: a disposition to deform in a certain way only under high strain. Yet it also has a power in another sense: it has the power to smash wineglasses, or ceramic cups, under certain conditions. A ceramic cup might also have a power in this sense in relation to a wineglass. In this sense of ‘power,’ a hammer is an *Aristotelian agent*. Something (e.g., a steel hammer) is an Aristotelian agent if whenever it acts on a patient (e.g., a wineglass) the change occurs in the patient.³⁷

³⁵ Cf. van Inwagen (1983: 8).

³⁶ If I respond by saying, “Oh, can’t I?” then I am predicating of myself an *ability*. One way to cash out the sort of permissibility talk discussed in the text along the lines of the possibility talk outlined earlier is to say that the full meaning of ‘S can (in the permissibility sense) A’ depends on the facts with which S’s A-ing is compossible, where these are moral or legal facts, or whatever apparently fact-like things make up the content of morality, law, rationality, etc. (Cf. Hobbs unpublished.)

³⁷ See e.g., Gill (1980).

There is a sense in which humans are agents that is distinct from the way in which hammers are agents. For instance, when we say that someone can speak Gaelic, we are ascribing to her a general ability—a power relating her to a potential action (speaking Gaelic is active, whereas understanding it is merely passive). Having an ability of any sort is to have what van Inwagen calls “a power to originate changes in the environment” (1983: 11). It seems that our *originating* such changes is (partly) what differentiates us as agents from hammers. We, unlike hammers, have abilities to *do* things, or to *make things happen*. On this taxonomy, it remains open whether powers or abilities just are, or are somehow analyzable in terms of, dispositions.³⁸ This also highlights a way in which it is different to attribute an ability to an agent than it is to say that something *can* happen. After all, even if I can be hit by a meteor, we do not want to say that I have the *ability* to be hit by a meteor.³⁹

General abilities are compatible with determinism: I retain the general ability to jump up and down on one foot at time *t*, without—in another sense—being able to exercise it just then, perhaps because I am asleep. Determinism is compatible with my having, at *t*, a general, but at that time unexercisable, ability to jump up and down on one foot.⁴⁰ Only specific abilities are at issue in debates about free will. In particular, we are interested in agents’ specific ability to A at *t*, or to refrain from A-ing at *t*.

Plausibly, general and specific abilities⁴¹ lie on a continuum, with the most general sort of ability at one end, and the specific ability to act and to refrain at the other. What happens in moving from the most general to the most specific ability is that the facts relevant to whether we actually possess the ability attributed to us become more narrowly specified. So, by analogy with David Lewis’s (1986)

³⁸ See e.g., Vihvelin (2004), Smith (2004), and Fara (2008), for optimistic views in this regard. See Clarke (2009) for a sobering appraisal of such optimism.

³⁹ I owe this point to James Hobbs (unpublished).

⁴⁰ Note that unexercised specific abilities are also compatible with determinism; see below.

⁴¹ See Honoré (1964) and Mele (2003) for somewhat different discussions of general and specific abilities.

contextualist account of ‘can,’ most generally it is true that I am able to speak Japanese, even if I do not know how to speak Japanese.⁴² Clearly, I am intrinsically constituted in such a way that I possess the properties that would allow me to speak Japanese in the right circumstances: nothing is preventing me from learning it. In this sense, I am able to speak Japanese, whereas an orangutan is not. This contrasts with a less (but still) general sense of ability. Assuming that I do not know how to speak Japanese, when I am among tourists and someone inquires of me whether I am able to speak Japanese, it is (as it happens) correct for me to reply, “No.” After all, I have not taken the time to learn it. Even having been called upon to speak Japanese, and despite my suitable intrinsic constitution, no matter how motivated I am to speak Japanese I still lack the specific ability to speak a word of that language.

We all possess both general and specific abilities to do things that we never do, and perhaps never will do. After all, even if I never learn how to speak Japanese, nevertheless I am still able (in the sense outlined above) to speak it. More specifically, if I am on holiday in Spain and someone offers me *criadillas* for lunch, I am able to eat them, although it is certain that I will refuse. Often, we try and fail to do things that we possess a specific ability to do. For instance, I am able to play “*Recuerdos de la Alhambra*” on guitar, and I possess the specific ability to do so right now. Yet, it is a difficult piece, and so I might try and fail on this occasion (Cf. Austin 1956).

What is a specific ability? We might think that my having a specific ability to play the guitar at a time, *t*, is for me (a) to have a general ability to do so, and

⁴² Lewis (1986) has a contextualist view of ‘can,’ according to which when we say that an agent, *S*, can perform an action, *A*, we are saying that *S*’s *A*-ing is compossible with certain facts. Here, whether it is true that *S* can *A* may depend on something left implicit—the set of facts with which *S*’s *A*-ing is compossible. Lewis’s view is contextualist since the meaning of ‘can’ does not, by itself, determine what facts are relevant; the additionally relevant facts are determined by context. So while all uses of ‘can’ share a semantic element—they express compatibility with certain facts—the precise meaning of particular use of ‘can’ depends on something else, which Lewis calls “context.” So, ‘*S* can *A*’ means that *S*’s *A*-ing is compatible with certain facts, where the relevant facts depend on the stringency with which ‘can’ is used. I discuss Lewis’s contextualist proposal in Chapter 3.

(b) to have an opportunity to exercise my general ability at t . Naturally, I retain the general ability to play the guitar even when I am miles away from one. If a guitar were brought before me at t , then I would gain an opportunity to exercise my general ability. Thus, I might possess the specific ability to play the guitar at t . Yet on this view, my specific ability to play the guitar at t and to refrain from playing the guitar at t turns out to be *obviously* compatible with determinism. In order for the view to account for the ability at issue in questions about whether modal freedom is compatible with determinism, we would need to say more about what it means to have an opportunity, and we cannot stipulate that having an opportunity is an incompatibilist notion, since doing so would beg the question against the compatibilist.

Perhaps, then, an agent, S, has a specific ability to A at t only if (i) S has a general ability to A at t ; (ii) S has an opportunity to A at t ; and (iii) holding fixed the past and the laws (including S's motivations regarding her opportunities as they are at t), S can exercise her general ability to A at t . If S has such an ability regarding two actions, A and B (where $A \neq B$), at t , then S has the specific ability to do otherwise at t . Here, (i) and (ii) are neutral on the compatibility issue. To locate the point of contention between compatibilist and incompatibilist, we add that even holding fixed S's motivations regarding her opportunities as they are at t , S can exercise her general ability to A at t and her general ability to refrain from A-ing at t by B-ing instead. Compatibilists think that, assuming determinism, S is free to do otherwise. Incompatibilists think that S is free to do otherwise only if determinism is false.

How are we to think about this disagreement between compatibilist and incompatibilist? After all, one might think it obvious that if determinism is true and there is only one physically possible future, then no one is free to do otherwise than they do. Accounting for how agents are able to do otherwise in a deterministic world *is* one of the deepest challenges that compatibilists face. There have been many proposals, and as many failures. Before mentioning some of these,

it may be helpful to lay a little groundwork in order to show what such a theory would have to look like.

To begin with, Alfred Mele (2003; 2006: 17–25) introduces a useful distinction between “simple abilities” (“S-abilities”) and “intentional abilities” (“I-abilities”), each of which may be general or specific, depending on whether one has the opportunity to perform the action in question. I will restrict discussion to specific abilities. I might have the S-ability to roll a six when I roll a die, in a sense in which I never have an S-ability to lift a bus, for instance. My having once thrown a six is, as Mele puts it (2006: 18), conceptually sufficient for my having the S-ability to throw a six right now. Yet it is not conceptually necessary that one already have done something in order to possess the S-ability to do it, since even an isolated native of the Amazon rainforest who is given a die for the very first time presumably also has the S-ability to throw a six. Thus, an agent *S* has a specific S-ability to perform an action *A* at a time *t* iff there is a possible world in which *S* A-s at *t*.⁴³

By contrast, I possess the I-ability to roll a die right now as long as I have a die to hand, but not the I-ability to roll a six. After all, I am not plausibly able intentionally to roll a six, and I do not intentionally roll a six just because on some occasion I intended to do so and I got lucky and did. So an agent *S* has a specific I-ability to *A* at *t* iff there is a relevant possible world in which *S* intentionally A-s at *t* (where this entails that *S* has an S-ability to *A*, but is not entailed by *S*’s possessing an S-ability). The idea is that we “graft” onto an account of a specific S-ability an account of intentional action, so that *S*’s A-ing also counts as *S*’s exercising an I-ability.

In trying to understand the disagreement between compatibilists and incompatibilists about the specific ability to do otherwise, we might ask what worlds we should consider when assessing whether *S* has an I-ability to perform more than one action at a time. As Mele puts it, “One way to see the disagreement

⁴³ Likewise, *S* has a general S-ability at *t* to *A* at *t* iff there is a possible world in which *S* A-s at some time.

between incompatibilists and compatibilists about determinism and being able to do otherwise is as a disagreement about what worlds are relevant” (2006: 21). For incompatibilists, “all and only worlds with the same past and natural laws as W [the actual world] are relevant; they hold the past and the laws fixed” (21). For compatibilists, by contrast, worlds with different pasts and natural laws than the actual world are relevant to whether S is able to perform or refrain from performing a given action at time *t*. Thus, assuming that in the actual world W an agent S does not perform some action A at time *t*, Mele suggests that we might define an intentional *libertarian* ability (L-ability) in the following way: An agent S has, at *t*, the L-ability to A intentionally at *t* in W, the actual world, iff there is a possible world with the same past and laws as W in which S A-s intentionally at *t* (Cf. Mele 2006: 19).

Compatibilists think that there is a sense in which agents are able to do otherwise that is consistent with determinism. They think that worlds with different pasts and laws than the actual world are relevant to judging whether agents have the ability to do otherwise at a given time in the actual world. Their accounts differ mainly in the way that they go about making use of such worlds. For instance, David Lewis (1981) argues if S A-s at *t*, nevertheless S had the specific ability to refrain from A-ing at *t* in the actual world, as long as there is a possible world in which S’s not A-ing at *t* is permitted by a small local exception to the laws. This is the “local miracle” view. On the “backtracking” view, if S A-s at *t*, nevertheless S had the specific ability to refrain from A-ing at *t*, as long as there is a possible world in which the laws are the actual laws, but events in that world prior to S’s A-ing are sufficiently different from those of the actual world to allow for S’s refraining from A-ing (Bennett 1984; Cf. Saunders 1968). For advocates of the compatibilist “conditional analysis” (e.g., Moore 1912), S is able to A and to refrain from A-ing at *t* iff were S to choose to A at *t*, S would A at *t*, and were S to choose to refrain from A-ing at *t*, S would refrain from A-ing at *t*.⁴⁴ Some

⁴⁴ This analysis fails, since the following conditional might be true: If S chose to refrain from A-ing, then S would refrain from A-ing. Yet, S might be unable to refrain from A-ing because she is unable to choose to refrain (Lehrer 1968: 32). There are recent revivals of this conditional

compatibilists understand the ability to do otherwise less stringently, by insisting that S is able to A and to refrain from A-ing at *t*, even if, holding fixed the laws and the state of the world at *t*, S is unable to exercise her general ability to refrain from A-ing at *t*.

One might worry that if we characterize the specific ability to do otherwise as I did earlier, then it is difficult to see how compatibilists and incompatibilists are disagreeing. Instead, they seem merely to be talking past one another. Recall, I suggested that an agent S has a specific ability to A at *t* only if (i) S has a general ability to A at *t*; (ii) S has an opportunity to A at *t*; and (iii) holding fixed the past and the laws (including S's motivations regarding her opportunities as they are at *t*), S can exercise her general ability to A at *t*. If S has an ability in this sense regarding two distinct actions at *t*, then S has the specific ability to do otherwise. The point of contention between compatibilists and incompatibilists is meant to be (iii). Yet, if compatibilists think that worlds with different pasts and laws than the actual world are relevant to whether agents are able to do otherwise at a given time in the actual world, but incompatibilists deny this, then each party is talking past the other.

One way to sharpen this worry is to use Angelika Kratzer's (1977) univocal semantics for the modal terms 'can' and 'must.' This approach is similar to Lewis's (1986) contextualist account of 'can,' mentioned earlier. According to Kratzer, 'can' plays a similar role wherever it appears. She captures this similarity by univocally treating 'can' as an existential quantifier over worlds that are restricted by a contextual "in-view-of" clause.⁴⁵ Thus, when I say, "You cannot move the pawn three spaces ahead," what I mean is, "In view of the rules of chess, you cannot move the pawn three spaces ahead." That is true. However, once the contextually restricting "in-view-of" clause is altered, what I say might be false. For instance, if I mean, "In view of your physical constitution, you cannot move the

or dispositionalist view (e.g., Vihvelin 2004; Fara 2008). See also Lehrer 1976 for a different compatibilist view.

⁴⁵ 'Must,' meanwhile, is treated by Kratzer as a universal quantifier over a contextually restricted set of worlds.

pawn three spaces ahead,” presumably that is false. A view of this sort spelled out in terms of one’s being able to do otherwise at a time could also be made univocally to treat ‘is able to’ as an existential quantifier over worlds restricted by the following “in-view-of” clause: “In view of the actual past and the actual laws, S is able to A at t or to refrain from A-ing at t by B-ing instead.” One might then insist that unless the compatibilist and incompatibilist agree on this restriction on the set of worlds to focus on, they are not treating ‘is able to’ univocally, and so are not disagreeing.⁴⁶

The strongest compatibilist position about the ability to do otherwise is one on which the compatibilist agrees to this contextual restriction on the set of worlds that are relevant. In the chapters that follow, I tell a compatibilist story about the experience of being able to do otherwise, and I will suggest a compatibilist account of such an ability that from the outset agrees that this is sometimes an appropriate sort of restriction to have in place when assessing claims about the ability to do otherwise.⁴⁷

For now, however, let me return to the natural compatibility question, which is a descriptive question about people’s actual belief tendencies regarding their freedom. In what follows, I suggest that *agentive experience* is relevant to addressing this question.

1.4 The Experience of Freedom

According to the way in which I have suggested we should address the question whether modal freedom is compatible with determinism, we must first ascertain what people’s belief-tendencies are regarding freedom. In other words, before we can address

⁴⁶ I am indebted to Alex Grzankowski for suggesting this way of putting things.

⁴⁷ See Chapter 5 (section 5.5.1) for details about how my compatibilist view differs from more traditional compatibilist views.

The freedom-to-do-otherwise-compatibility question: Is the freedom to do otherwise compatible with determinism?

we must first answer

The freedom-to-do-otherwise-descriptive question: What sort of freedom do people tend to believe they have?

The compatibility question we are asking here is whether people are free in the way that they tend to believe they are, assuming determinism.⁴⁸ This is one of the central questions that grips people when they first consider their agency in relation to determinism. After all, when people first entertain the compatibility question, they do not think about the transfer-of-powerlessness principle that lies at the center of debate about the consequence argument, or the complex Frankfurt-cases at the center of debates about whether responsibility requires alternative possibilities (Cf. Frankfurt 1969). Rather, they are struck by the thought that determinism robs them of their freedom to do otherwise. (Or so I argue in Chapter 2.) People are struck by the fact that determinism implies that there is only a single physically possible future: “If *that’s* right,” people think, “then I’m not free in the way that I thought I was.”

The crucial question, then, is the freedom-to-do-otherwise-descriptive question, on which the answer to the compatibility question depends. Furthermore, the freedom-to-do-otherwise-descriptive question is an empirical question. If people tend to believe that they possess an incompatibilist freedom to do otherwise, then the answer to the compatibility question will be that people are natural incompatibilists about such freedom. It will be a further question, of course, whether people are *actually* libertarian agents. That, too, is an empirical question, the answer to which depends on what our best scientific theories say about the nature of the universe and human decision-making. This is what I earlier called the substantive question. I leave this question aside as work for another day. I assume that it is

⁴⁸ Or “near determinism”; see section 1.2 above.

unlikely that people are libertarian agents, and I further assume (at least for argument's sake) that determinism is true. Thus, I address only the descriptive and prescriptive questions, and only in relation to the question whether the freedom to do otherwise is compatibilist.

How should we address the descriptive question? One way is simply to go out and empirically measure people's belief-tendencies about the freedom to do otherwise. In other words, we might gauge people's intuitions about being free to do otherwise. Some empirical work has already been done on this issue by experimental philosophers, although with inconclusive results (e.g., Nichols 2004; Turner and Nahmias 2006).

I noted earlier that some philosophers think that our being free to do otherwise is confirmed by our experience of deliberation and choice. In section 1.3, I noted that incompatibilists have long appealed to this sort of experience in order to explain belief in libertarian free will, and also to justify such beliefs. Most compatibilists, of course, deny such claims. Instead, they claim that our experience carries with it only compatibilist commitments. This suggests another way in which we might gauge whether people's commitments about freedom are compatibilist: try to find out whether people's *experience* of freedom is veridical, assuming determinism:

The experience-compatibility question: Is people's experience of being free to do otherwise veridical, assuming determinism?

Just as with the freedom-to-do-otherswise compatibility question, the answer to the experience-compatibility question depends on our answer to a related descriptive question:

The experience-descriptive question: What sort of freedom to do otherwise do people experience having?

What I am proposing is that experiences are as legitimate a starting point as intuitions are for investigating people's commitments about freedom. Leaving aside

questions about the justificatory status of intuitions, intuitions appear at least to be reliable indicators of how things *seem to be* to a person (Cf. Bealer 1998). Furthermore, having the intuition *that p* often motivates one (to some degree) to accept the content *p*.⁴⁹ Thus, one may have a tendency of some strength to believe that *p*. By doing empirical work on the freedom-to-do-otherwise-descriptive question, philosophers treat intuitions as reliable indicators of how things seem to be to people, regarding their modal freedom. Is there another reliable indicator of this sort? Yes, there is. Another way in which things might seem to be a certain way is experientially. For instance, a stick in the water might seem to be bent: this seeming is experiential. So, there are different ways that something might seem to be the case. As a result, there are different ways in which one might be motivated to believe that a content, *p*, is veridical. Intuitions (as “intellectual seeming-states”) are one way, while experiences (“experiential seeming-states”) are another.⁵⁰

I will assume that people’s introspective reports about their experience can be taken at face value. In assuming this, I follow Uriah Kriegel (forthcoming) in steering a course between “introspective dogmatism,” according to which introspection is infallible, and “introspective skepticism,” according to which introspection is utterly unreliable. Kriegel defends “introspective moderation,” which is the view that introspection has *above-chance* reliability. Kriegel thinks that introspection is at least as trustworthy as a normal adult’s sense of smell, for instance. After all, he says, my smelling raspberries makes it more probable that there are in fact raspberries nearby than if I do not smell them; and when there are actually raspberries nearby, it is more likely that I will smell them than if there are none. Likewise, my introspecting a certain phenomenology, *P*, makes it more

⁴⁹ This motivation may, in some cases, be based on one’s relative competence in an area. See Symons (2008) and Bedke (2008) for useful discussions of intuitions in philosophy. See also Jenkins (forthcoming).

⁵⁰ Here, I am deliberately running together both perceptual and introspective “seeming” states. There is much to be said (and that has been said) about the verb ‘to seem’ and other “appear words” (see, e.g., Chisholm 1957; Brogaard 2013). However, these issues are orthogonal to my present concerns.

likely that I have P than if I do not introspect it; and my having P makes it more likely that I will indeed introspect P than if I do not have P. Of course, for introspection (as for smell), this claim about reliability may hold only for normal subjects under normal conditions. It might not hold for elusive phenomenologies. For instance, there may be “fringe” phenomenologies, such as my peripheral sense of my legs pressing against my chair. When focused upon, this phenomenology loses its fringe quality. Thus, it seems not to be introspectible, since it disappears as soon as I turn my introspective attention upon it. Yet, once we grant these caveats, it seems plausible that introspection does indeed have above-chance reliability.

If someone introspectively reports having an experience as of being free to do otherwise that has libertarian content, the default position that I will adopt is to grant that they indeed have such phenomenology. If experiential seeming-states drive belief in much the same way that intuitions do, then someone’s introspective report about her experience of being free to do otherwise is roughly as reliable an indicator of her being disposed to believe that she possesses such freedom as the relevant intuition would be. In other words, if people report having experiences that are libertarian in nature, then, in the absence of countervailing considerations, we may assume that people actually do tend to believe that they possess libertarian freedom.⁵¹

One item stands in need of clarification. When we ask whether someone’s experience as of being free to do otherwise is compatible with determinism,⁵² we are asking whether her experience is *veridical*, assuming determinism. We are not asking whether the experience itself is compatible with determinism, even if it has libertarian content. After all, presumably any experience whatever, including

⁵¹ I discuss these issues further in the chapters that follow.

⁵² I will not always adhere to the locution “the experience *as of* being able to otherwise.” Often, for ease of exposition, I will simply say “the experience *of* being able to otherwise.” Unless otherwise indicated, the latter phrase should be read in the former way. In other words, experiences can be in error.

a libertarian one, is compatible with determinism, given that we could be determined to have that very experience.⁵³

1.5 Outline of the Project

Here is how I will proceed. In Chapter 2, “Phenomenal Abilities: Incompatibilism and the Experience of Agency,” I present the results of a series of studies that I conducted together with Matt Bedke and Shaun Nichols (2013), which indicate an incompatibilist answer to the experience-descriptive question. We found that participants in our studies tended to report having an incompatibilist experience of being free to do otherwise. In Chapter 3, “Is Agentive Experience Compatible with Determinism?” I anticipate a compatibilist objection to these results; namely, that people’s incompatibilist reports about their experience are prone to error in a particular way. I defend the results described in Chapter 2 against an objection of this sort developed by Terry Horgan (e.g., 2011). In place of Horgan’s compatibilist view, I propose an alternative prescriptive strategy for compatibilists, which I outline at the end of Chapter 3 and develop at length in Chapter 4, “Against an Argument for Libertarianism.” In Chapter 5, “Indeterminism, Experience, and Compatibilism,” I explain how people get to have indeterminist experiences of freedom, and thereby incompatibilist beliefs about their freedom to do otherwise. Yet, the account I offer falls short of justifying belief in libertarianism. On the contrary, I explain how such experience is consistent with compatibilism about modal freedom, and thus I defend compatibilism about being free to do otherwise.⁵⁴ In Chapter 6, I review the dissertation, showing how its various strands hang together.

⁵³ Mele (1995: 133–37, 246–49) makes this point.

⁵⁴ Chapters 2–5 were prepared in the first instance for submission to refereed journals. Even so, these chapters (together with the introductory and closing chapters) form a coherent document that provides a systematic account of my research project, as per FoGS requirements at UBC. The development of my argument throughout the following chapters is presented, in form and content, as a unified whole.

Chapter 2

Phenomenal Abilities: Incompatibilism and the Experience of Agency

Incompatibilists often claim that we experience our agency as incompatible with determinism, while compatibilists challenge this claim. In this chapter, I report a series of experiments that I conducted together with my colleagues Matt Bedke and Shaun Nichols. These experiments focus on whether the experience of having an ability to do otherwise is taken to be at odds with determinism. Participants in these studies described their experience as incompatibilist whether the decision was (i) present-focused or retrospective, (ii) imagined or actual, (iii) morally salient or morally neutral. The only case in which participants did not give incompatibilist judgments was when the question was explicitly about whether one's ignorance of the future was compatible with determinism. This lends empirical support to claims made by incompatibilists about the experience of agency, while also showing that compatibilist accounts of ability are inadequate to the reported

phenomenology. These results also inform recent debates about the presuppositions of deliberation.¹

2.1 Background

2.1.1 Introduction

Agents act. They buy detergent at the store, they go to work, they celebrate holidays, they cheat on their taxes. Sometimes we hold agents morally responsible for what they do, or what they fail to do, meting out credit and blame as the occasion merits. In typical cases, when agents act they are thought to have an *ability to do otherwise*. This is a point on which most parties to the free-will debates agree. When it comes to characterizing the ability to do otherwise and asking whether this ability is compatible with determinism, however, there is no consensus.

In the ensuing debates, the *experience* as of having an ability to do otherwise occupies a central role.² Many libertarians, for instance, maintain that the ability experienced is incompatible with determinism (C. A. Campbell 1951; O'Connor 1995). Of course, some compatibilists have challenged this idea (Mill 1865; Grünbaum 1952; Nahmias et al. 2004). Despite the centrality of the phenomenol-

¹ At the time of submission of this dissertation, Chapter 2 has just been published as: Oisín Deery, Matt Bedke, and Shaun Nichols (2013), “Phenomal Abilities: Incompatibilism and the Experience of Agency,” in D. Shoemaker (ed.), *Oxford Studies in Agency and Responsibility: Vol. 1*, 126–150. It is reprinted here with the permission of Oxford University Press (UK), and has been altered to maintain consistency with the rest of this dissertation (e.g., additional material has been added at the beginning and end of the chapter).

Thanks to Eddy Nahmias, Dana Nelkin, Paulo Sousa, David Shoemaker, Uriah Kriegel, and Bertram Malle for comments on earlier drafts. Many thanks also to participants at the New Orleans Workshop on Agency and Responsibility (November 3-5, 2011) and two anonymous reviewers at OUP for suggestions and comments.

² For some of the literature on the ability to do otherwise, see Moore (1912), Berofsky (2002), Joseph Campbell (2005), Perry (2004), Vihvelin (2004), Smith (2004), Fara (2008), Fischer (2008), Clarke (2009). The claim that moral responsibility requires alternative possibilities has been disputed since Frankfurt (1969). However, it is still widely contended that free will requires being able to do otherwise.

ogy of agency in all this, there has been strikingly little work on its characteristics. Of particular significance, there is almost no empirical work on whether the experience of agency involves a phenomenology of being able to choose among alternative possibilities or whether people take their agentive experiences to have incompatibilist elements.³

This paper reports a series of experiments that investigates the phenomenology of agency. To anticipate, my colleagues and I found remarkably consistent results across three sets of studies: participants regarded their experience of the ability to do otherwise as incompatible with determinism. Now that any suspense has been spoiled, I will locate the issue in the broader literature.

2.1.2 The Experience of the Ability to Do Otherwise

Let us characterize determinism as follows: a statement of the facts of the world at an instant together with a statement of the laws of nature entail all truths about the world, including those about future human actions.⁴ Granting that we often feel that we have an ability to act other than we do, for present purposes incompatibilists think that the experience as of an ability to do otherwise is incompatible with determinism, while compatibilists think the opposite.

There are a number of influential introspectors on both sides of this issue. John Searle is a representative incompatibilist:

[R]eflect very carefully on the character of the experiences you have as you engage in normal, everyday human actions. You will sense the possibility of alternative courses of action built into these experiences. . . that we could be doing something else right here and now, that is, all other conditions remaining the same. This, I submit, is

³ One exception in the recent literature is a paper by Nahmias and colleagues (2004), which I discuss in section 2.1.3. Although my colleagues and I challenge their experimental results, we are indebted to them for pioneering the investigation. I also draw on their scholarship in setting out some of the historical statements below. See also Monroe and Malle (2010).

⁴ In the experiments below, my colleagues and I present this idea in terms of causation to make it more intuitive and accessible.

the source of our own unshakeable conviction of our own free will.
(1984: 95)

Similarly, Keith Lehrer has claimed that the incompatibilist “accurately describes what I find by introspecting, and I cannot believe that others do not find the same” (1960: 150). Even such a paradigmatic compatibilist as David Hume (1960/1739) agrees with this sentiment when he writes, “There is a false . . . experience . . . of the liberty of indifference” (Bk. II, Part III, II).

The appeal to an incompatibilist phenomenology plays a particularly important role in libertarianism. Many libertarians maintain both that we experience our agency as incompatible with determinism, and that this experience provides reason to think that our agency defies determinism. C. A. Campbell writes:

Everyone must make the introspective experiment for himself: but I may perhaps venture to report. . . that I cannot help believing that it lies with me here and now, quite absolutely, which of two genuinely open possibilities I adopt. (1951: 463)

Campbell goes on to argue that, unless we have good reason to doubt the impression that “it lies with me” which of two possibilities I adopt, we should accept the impression to reflect the truth. Timothy O’Connor makes this move as well. First, O’Connor describes the character of the experience of decision-making:

[T]he agency theory is appealing because it captures the way we experience our own activity. It does not seem to me (at least ordinarily) that I am caused to act by the reasons which favor doing so; it seems to be the case, rather, that *I* produce my decision *in view of* those reasons, and could have, in an unconditional sense, decided differently. . . (1995: 196)

Next, O’Connor says that we should take these experiences to reflect something important about the nature of decision-making:

Such experiences could, of course, be wholly illusory, but do we not properly assume, in the absence of strong countervailing reasons, that things are pretty much the way they appear to us? ... Skepticism about the veridicality of such experiences has numerous isomorphs that, if accepted, appear to lead to a greatly diminished assessment of our knowledge of the world, an assessment that most philosophers would resist. (1995: 196–197)

A number of compatibilists have challenged the basic phenomenological claim. These compatibilists deny that we experience our agency as incompatible with determinism. John Stuart Mill, for instance, writes,

Take any alternative: say to murder or not to murder... If I elect to abstain: in what sense am I conscious that I could have elected to commit the crime? Only if I had desired to commit it with a desire stronger than my horror of murder; not with one less strong. When we think of ourselves hypothetically as having acted otherwise than we did, we always suppose a difference in the antecedents: we picture ourselves as having known something that we did not know, or not known something that we did know; which is a difference in the external motives; or as having desired something, or disliked something, more or less than we did; which is a difference in the internal motives. (1865: 285)

On Mill's view, the feeling of the ability to do otherwise is always contingent on our supposing that the situation prior to the decision was somehow different. Adolf Grünbaum repudiates any incompatibilist element with equal vigor:

Let us carefully examine the content of the feeling that on a certain occasion we could have acted other than the way we did... Does the feeling we have inform us that we could have acted otherwise under exactly the same external and internal motivational conditions? No,

... this feeling simply discloses that we were able to act in accord with our strongest desire at that time, and that we could indeed have acted otherwise if a different motive had prevailed at the time. (1952: 672)

Grünbaum's last sentence here gestures at the payoff of denying the phenomenological claim of incompatibilist agency: if Mill and Grünbaum are right, then the feeling of being able to do otherwise is consistent with determinism, and this would undercut a crucial motivation for libertarianism.

This situation might seem to be a dialectical stalemate (Cf. Fischer 1994: 84). However, these philosophers are making general claims about the nature of our experience of agency. These are empirical claims, and they can be illuminated by taking up empirical methods.

2.1.3 Previous Work on the Phenomenology of Free Will

My colleagues and I are not the first to recommend a more systematic investigation that is partly empirical. Nahmias and colleagues (2004) suggest that we find out how people actually tend to describe their agentic experience (what they call the phenomenology of free will), including their experience as of being able to do otherwise:

Taking a cue from recent empirical work on “folk intuitions,” we think the best way to understand the phenomenology of free will—if there is one—is to find out what ordinary people's experiences are like. If this is not possible, philosophers' competing introspective descriptions will remain in yet another free-will stalemate. (164)

Nahmias and colleagues undertook this task in survey studies. Their studies appear to lend some support to the idea that the phenomenology of agency is compatibilist. However, these studies have significant shortcomings, so let me briefly describe one of those studies, and then identify what is lacking.

In one study, Nahmias and colleagues pitted compatibilism and incompatibilism against each other directly. The study was based on “competing libertarian

and compatibilist accounts of our experience of the ability to choose otherwise” (174). Their survey asked participants to imagine (or recall) an experience of making a difficult choice:

Imagine you’ve made a tough decision between two alternatives. You’ve chosen one of them and you think to yourself, “I could have chosen otherwise” (it may help if you can remember a particular example of such a decision you’ve recently made).

Which of these statements best describes what you have in mind when you think, “I could have chosen otherwise”?

- A. “I could have chosen to do otherwise even if everything at the moment of choice had been exactly the same.”
- B. “I could have chosen to do otherwise only if something had been different (for instance, different considerations had come to mind as I deliberated or I had experienced different desires at the time).”
- C. Neither of the above describes what I mean. (2004: 175–76)

The majority of the participants gave the response that fits with compatibilism (i.e., B).

While this study is clearly focused on an issue that divides compatibilists and incompatibilists, there are a number of limitations to the study. First, participants are told to think of a decision and then told to think something else about the decision: that they *could have done otherwise*. It is thus unclear whether their initial recollection *actually* carried with it a sense of an ability to do otherwise. So if people make compatibilist judgments about these decisions, it might be because they are considering cases in which the phenomenology of the ability to do otherwise is absent.

Second, the key question is about experiences sometime in the past, rather than present-focused experiences where the phenomenology of agency is actually present and thus presumably more accessible.

Third, Nahmias and colleagues asked participants about *difficult* decisions, and this presents the opportunity to interpret “could have done otherwise” in confounding ways. Consider Martin Luther’s decision to renounce his writings or be declared an outlaw and heretic. Legend has it that, after praying and consulting with advisors for a day, he said, “Here I stand. I can do no other,” thereby reaffirming his writings. Luther might have chosen B in Nahmias’s survey. But if he did, we should not conclude that there is no sense of ‘could have done otherwise’ that captures some aspect of Luther’s phenomenology and that is incompatible with determinism. For Luther could have responded as he did to express his commitment to his cause, a commitment that would only change if the considerations before him and his reasons for breaking with the Roman Catholic Church presented themselves differently. This commitment-expressive meaning of ‘could not have done otherwise’ is consistent with other senses of ‘could have done otherwise’—consider whether Luther thought it was up to him whether to renounce his views—that might or might not be incompatible with determinism. Difficult decisions are subject to confounds like this, so the above survey does not cleanly address the question whether there is some aspect of the phenomenology of agency that is in tension with determinism.

Fourth and last, it is not clear whether the participants really understand the intended meaning of “even if everything at the moment of choice had been exactly the same” or “only if something had been different.”

My colleagues and I wanted to run more comprehensive studies that fix these shortcomings. The result was the following three studies, which share a common structure. First, participants were asked whether they had an experience as of the ability to do otherwise when faced with a simple decision. Next, they were given a description of determinism. Of course, we did not use the term ‘determinism,’ since that might have conjured up unwanted associations in par-

ticipants. Rather, we used a technical term—‘causal completeness.’ To address concerns about comprehension of the materials, the familiar psychological technique of *training to criterion* was adopted, thus participants were asked a series of questions that tested and, if necessary, corrected, their understanding of determinism. Participants who passed the training were asked about the compatibility of their experience with determinism. In study 1, this question focused on both a first-person, present-focused experience in a hypothetical deliberative context and a past-focused judgment about such a situation. Study 2 explored the phenomenology of actual rather than imagined choices. And study 3 tested whether epistemic phenomenology—the phenomenology of uncertainty—feels incompatible with determinism.

2.2 Study 1

Overview

In the first study, Bedke, Nichols and I had participants imagine a decision about whether to go left or right on a sled. In one condition, the sledding scenario was set in the future; in the other condition, the scenario was set in the past. After reading the scenario, participants in condition 1 were asked whether they had a feeling of an ability to do otherwise; participants in condition 2 were asked for a retrospective judgment about whether they could have done otherwise. Participants who affirmed feeling (or having) an ability to do otherwise were directed to the training section in which causal completeness (i.e., determinism) was explained to them. Participants who passed the training were reminded of their affirmation regarding the ability to do otherwise and asked about consistency with causal completeness.

The prediction my colleagues and I made was that when asked about the phenomenology of imagined decision making, participants would tend to affirm a

feeling of an ability to do otherwise and also regard this feeling as incompatible with determinism; but when asked for a retrospective judgment about the ability to do otherwise, we predicted that participants would be less likely to treat the ability to do otherwise in an incompatibilist way.⁵

Method

Participants:

84 participants were initially recruited online through the Mechanical Turk (MTurk) website.⁶ The survey itself was conducted using SurveyMonkey. 2 participants did not complete the survey. They were excluded from the analysis.

Materials:

Each condition had three parts.

Part 1: The ability to do otherwise

Participants were presented with a vignette and a question about the ability to do otherwise. For condition 1, this went as follows:

Please read the following passage, and answer the questions that follow as best you can:

Imagine that you are sledding down a snowy path on a mountainside. Your sled has a steering mechanism that allows you to control the direction of the sled. Below you is a fork in the path with snow built up in the middle, and you can tell that, if you don't direct your sled

⁵ See e.g., van Inwagen 1983 (8–13) for an overview of other uses of 'can.'

⁶ MTurk is a website supported by Amazon.com (<https://requester.mturk.com/mturk/welcome>) that provides users the opportunity to fill out surveys for modest compensation. Recent work indicates that the data gathered through MTurk is at least as reliable as that gathered through standard psychology pools composed of undergraduates (see Buhrmester et al. 2011).

one way or the other, the contours of the mountain will channel you and your sled either to the left or to the right.

Ability Question

Consider how things seem to you as you approach the fork in the path. In particular, consider what it's like to decide which way the sled will go.

Please indicate your level of agreement with the following statement:

When deciding which way the sled will go, it feels like I can either go to the left or go to the right.

Participants were asked to indicate their agreement with this statement on a 7-point scale (1=disagree completely; 7=agree completely).

For condition 2, the vignette was the same except that participants were asked to imagine that the sledding episode occurred many years ago. And instead of getting a response regarding a phenomenology of the ability to do otherwise, we asked them to indicate agreement with a statement about a past ability to have done otherwise: "I could have gone right instead of left."

Part 2: Training on determinism

The aim was to focus on participants who had a phenomenology of the ability to do otherwise, so only participants who indicated a positive level of agreement to the first questions (5 or higher) were directed to the training section. Here, participants were given a detailed explanation of causal completeness, summed up as follows: "According to causal completeness, everything that happens is fully caused by what happened before it. This is true from the very beginning of the universe, so what happened in the beginning of the universe fully caused what happened next, and so on right up until the present. Causal completeness holds that everything is fully caused in this way, including people's decisions."

Participants were then given two kinds of cases to illustrate the phenomenon. In one case, they were told that an earthquake fully caused the volcanic eruption at Mt. St. Helens,⁷ and they were then told, “According to causal completeness, if we could somehow replay the entire past right up until St. Helens erupted on May 18, 1980, then St. Helens would once again erupt at that time. Another way to put this is to say that all the events leading up to the eruption made it so that the eruption had to happen.” In another case past events, feelings and beliefs led to Obama’s decision to pick Joe Biden as his running mate, and participants were told “According to causal completeness, if we replayed the past right up until Obama’s decision—including everything that was going through Obama’s mind—then Obama would once again make exactly the same decision. That is, all the events leading up to Obama’s decision (including everything that was going through Obama’s mind), made it so that it had to happen that Obama would pick Biden.”⁸

We then tested comprehension of causal completeness. First, we asked participants to indicate whether the following was true or false:

According to causal completeness, St. Helens would have erupted on May 18, 1980 even if there had been no earthquake.

Participants who answered “True” (the incorrect answer) were corrected, and given an explanation of the right answer. These participants were then given a similar question to see if they had absorbed the training. If they answered incorrectly yet again, they did not move on to answer the compatibility question, as they were deemed to have insufficient comprehension of causal completeness.

⁷ This is an oversimplification of the geological facts which my colleagues and I adopted to ease the load on participants.

⁸ In defining determinism—our causal completeness (CC)—as meaning “everything that happens is *fully* caused by what happened before it,” some might think this consistent with certain indeterminist conceptions of causation. But to say that events are fully caused is meant to avoid this reading—being fully caused suggests that nothing extra-causal is needed to help settle events. My examples aid the preferred interpretation. I say, e.g., that under CC, “it had to happen that Obama would pick Biden.”

Participants who passed this first kind of question either on the first or second try were given another true/false question to test for comprehension:

According to causal completeness, if a week from now Barack Obama decides to have soda with dinner, all the events leading up to that decision will make it the case that he has to decide to have a soda with dinner.

The objective here was to test for and correct overly weak interpretations of causal completeness. Those who answered “False” (the incorrect answer) were corrected and given another chance at a similar question. If they answered incorrectly yet again they failed the training and did not answer the compatibility question. Participants who passed both kinds of questions either on the first or second try were deemed to have adequate comprehension of determinism, and these participants moved on to the third part of the study, the compatibility question.

Part 3: Consistency

After successful completion of the training, in both conditions participants were told to recall their agreement with the statement regarding the ability to do otherwise (from Part 1 of the survey). E.g., in condition 1, they were told:

Now, recall that you previously agreed with the following statement:

When deciding which way the sled will go, it feels like I can either go to the left or go to the right.

Following this, they were asked the compatibility question. In condition 1, this read as follows:

Compatibility Question

Considering this previous statement and your understanding of causal completeness, please indicate your level of agreement with the following:

Even though it felt like I could either go to the left or go to the right, if causal completeness is true there is something mistaken about how that decision felt to me.

In condition 2, the compatibility statement was:

Even though I said I could have gone right instead of left, if causal completeness is true there is something mistaken about what I said.

Agreement was indicated on the same 7-point Likert scale as was used for the Ability Question, and an answer above 4 was taken to be an incompatibilist answer.

Results

Of the 34 participants who started condition 1, 33 completed it. Of these, 31 indicated a phenomenology of an ability to choose among possibilities and all but 4 of them passed the training section.⁹ The remaining 27 participants gave a mean response of 4.93 on the compatibility question, which differed significantly from the midpoint of the scale, $t(26) = 2.65$, $p = .014$. That is, participants tended to interpret their agentic experience as being incompatible with determinism.

In condition 2, of the 50 participants who started the survey, 49 completed it. Of these, 47 indicated an ability to choose among possibilities and all but 2 of them passed the training section.¹⁰ The remaining 45 participants gave a mean response of 5.24 on the compatibility question, which differed significantly from the midpoint of the scale, $t(44) = 5.05$, $p < .001$.

A t -test comparing conditions 1 and 2 showed no significant difference, $p = .448$. So participants tended to be just as incompatibilist about retrospective

⁹ Nine participants required correction, and ultimately passed the training section. The responses of those who required correction did not differ from those who answered correctly without training ($p > .2$).

¹⁰ Seven participants required correction, and ultimately passed the training section. Again, there were no differences between those who required correction and those who didn't.

judgments of their ability to do otherwise as they are about their current experience as of being able to do otherwise. This first study provides evidence that people do indeed judge that their experience of deciding is inconsistent with determinism, in the sense that the experience is somehow mistaken or non-veridical if determinism is true. It also suggests that the effect is robust across retrospective and present-focused cases.

2.3 Study 2

Overview

One major limitation of study 1 is that it involved merely imagined choices. This inserts a distance between the actual phenomenology of decision-making and judgments about that phenomenology. As a result, study 2 introduced conditions in which agents actually make decisions. In addition, study 1 focused on decisions that have no moral weight. My colleagues and I thus added a condition in study 2 in which the decision *does* have a moral element. So this study comprises three conditions to test for any effect from actual choices or from morally salient choices. We also introduced two innovations to the study's design.

First, the vignette was more “choicey”. It struck us that in situations such as sledding down a hill often we don't have a salient experience as of deciding which way to go. We just go one way or the other. Second, we wanted to address a potential worry about the use of the word “mistaken” in the compatibility question. For example, condition 1 from the first study asked:

Even though it felt like I could either go to the left or go to the right, if causal completeness is true there is something mistaken about how that decision felt to me. (Emphasis added.)

One worry about this wording was that participants might misinterpret it as asking whether they were mistaken in thinking that their experiences felt a certain

way, rather than as asking whether there would be something mistaken about the content of their felt experiences.¹¹ The solution was to replace the above wording with a wording of the following form:

Even though it felt like I could either choose to X or choose to Y, if causal completeness is true then I couldn't really have chosen differently than I did.¹²

With these modifications, condition 1 presented participants with an *imagined* choice among two very similar charities, condition 2 presented participants with an *actual* choice among two similar charities, and condition 3 presented participants with an *actual morally salient* choice among two charities, one for endangered trees, another for children's cancer treatments.

Method

Participants:

155 participants were initially recruited online through the Mechanical Turk (MTurk) website. The survey itself was conducted using SurveyMonkey. 21 participants did not complete the survey, or indicated that they had recently taken a "very similar" survey.¹³ They were excluded from the analysis.

¹¹ Thanks to Lucas Thorpe for this objection.

¹² Two reviewers worried that, with causal completeness in mind (earlier described in terms of events that "had to happen") participants fix on one reading of the modal "couldn't really have chosen differently" and on that reading they give an "incompatibilist" response, whereas the description of their phenomenology might invoke a different reading of the modal that would not merit an incompatibility response. Of course, the key issue is whether participants feel their phenomenology wouldn't be veridical if CC were true. The question in study 3 is designed to first refer to the participants' reports on their phenomenology—that it felt like they could choose X or choose—and this helps subjects to focus on *that* modal content and whether *it* would be veridical if CC were true. Further, study 1 asked the compatibility question using different language that avoids this worry. My colleagues and I get the same incompatibilist results there.

¹³ Studies 2 and 3 were run after study 1, and some of the conditions in studies 2 and 3 were run serially, so we excluded participants who indicated they had taken a very similar survey to minimize the influence of having previously taken one of our surveys.

Materials:

As in study 1, this study had three parts.

Part 1: The ability to do otherwise

For condition 1 of this study, I asked participants to imagine deciding between two charities for endangered trees.

Imagine that you have \$0.50 to donate. You have two options:

Donate to a foundation that protects the endangered tree *Castanea Dentata*.

OR

Donate to a foundation that protects the endangered tree *Ulmus Dentata*.¹⁴

These are your only two options.

Condition 2 was similar except that participants were given an actual choice. Participants were told that they had \$0.50 to donate to one of the two tree charities. Participants were informed (truly) that we would actually donate this money to whichever charity they chose. Participants read:

You have \$0.50 to donate. We, the researchers, will actually donate this money for you whichever way you decide.

¹⁴*Castanea Dentata* and *Ulmus Dentata* are the names of the American Chestnut and the American Elm, respectively. They are endangered species in North America. The charities used were The American Chestnut Foundation (<http://www.acf.org/>), and Trees Winnipeg: Coalition to Save the Elms (<http://www.savetheelms.mb.ca/>).

Participants were then presented with the same option language as in the imagined condition, and each option appeared as a radio button at the bottom of the page.

Finally, condition 3 presented participants with a morally salient choice between a foundation that protects the tree *Castanea Dentata* or (2) and The Childhood Cancer Foundation,¹⁵ on the assumption that people tend to think that saving dying children has greater moral weight than saving endangered trees.

In all conditions, after being given the imagined or actual choice, participants were asked a question about the ability to do otherwise. For instance, in condition 3, participants were asked to indicate their level of agreement (on a 7-point scale) with the following statement:

When deciding which option to choose, it feels like I can either choose to donate to the endangered tree *Castanea Dentata* or choose to donate to the Childhood Cancer Foundation.¹⁶

(In conditions 2 and 3, participants were subsequently required to make a choice between the charities.) As in study 1, only participants who agreed with the ability-to-do-otherwise statement proceeded to the training.

Part 2: Training on determinism

The training section was the same as that used in study 1, and once again only those who passed the training proceeded to the compatibility question.

Part 3: Consistency

The compatibility question was adapted for the new cases. For example, in conditions 1 and 2, participants were asked to indicate agreement (on a 7-point scale) with this statement:

¹⁵ <http://www.candlelighters.ca/>

¹⁶ There is a potential concern here about this phrasing, which must be forestalled. Suppose I am determined to choose p. It follows that I can choose p. And you might think it further follows that I can choose p or choose q, for this follows from the simple logical principle of disjunction introduction. In that case the ability to choose p or choose q is clearly compatible with determinism. However, participants report an incompatibilist phenomenology as of an ability to choose p or q, which suggests that they are not reading “can choose p or can choose q” in this compatibilist way.

Even though it felt like I could either choose to donate to *Castanea Dentata* or choose to donate to *Ulmus Dentata*, if causal completeness is true then I couldn't really have chosen differently than I did.

Results

Of the 50 participants who started condition 1, 42 completed it and had not recently taken a very similar survey (3 had). Of these, 38 indicated a phenomenology of an ability to choose among possibilities and all but 3 of them passed the training section. The remaining 35 participants gave a mean response of 5.60 on the compatibility question, which differed significantly from the midpoint of the scale, $t(34) = 6.08$, $p < .001$. The results of an imagined choice are consistent with the results of condition 1, study 1, if not stronger by virtue of the more “choicely” vignette.

In condition 2, of the 48 participants who started the survey, 42 completed it and had not recently taken a very similar survey (4 had). Of these, 39 indicated a phenomenology of an ability to choose among possibilities and all but 2 of them passed the training section.¹⁷ The remaining 37 participants gave a mean response of 5.78 on the compatibility question, which differed significantly from the midpoint of the scale, $t(36) = 6.85$, $p < .001$. That is, participants were again incompatibilist about the phenomenology, this time of an actual choice.

In condition 3, of the 57 participants who started the survey, 50 completed it and had not recently taken a very similar survey (3 had). Of these, 43 indicated a phenomenology of an ability to choose among possibilities and all but 3 of them passed the training section.¹⁸ Most of the remaining 40 participants (90%) opted

¹⁷ Six participants required correction and successfully passed the training section. There was no difference between the responses of those who required correction and those who didn't ($p > .2$).

¹⁸ Six participants required correction and successfully passed the training section. Again, I found no difference between the responses of those who required correction and those who didn't ($p > .2$).

to donate to the Childhood Cancer Foundation, as expected on the assumption that the cancer charity would be regarded as more morally salient. The 40 participants gave a mean response of 5.85 on the compatibility question, which differed significantly from the midpoint of the scale, $t(39) = 7.66$, $p < .001$. Once again, participants reported incompatibilist phenomenology—this time with a morally salient choice.

ANOVA testing showed no overall effect of condition among conditions 1, 2, and 3, $F(2, 111) = .254$, $p = .776$. So there appears to be no effect produced by making the condition an actual choice, or by making the choice morally salient.

The results of study 2 show that people report incompatibilist phenomenology of agency for actual choices. Indeed, whether the decision is set up as an imagined one or an actual one does not affect the degree to which participants interpret their agentic experience as being incompatible with determinism. The results also show that whether or not the decision is morally salient doesn't affect the degree to which participants interpret their agentic experience as being incompatible with determinism. So the results of previous studies seem to extend to the moral domain, where issues of responsibility loom large.

2.4 Study 3

Overview

One possible concern with the previous studies stems from the way in which the key compatibility question is phrased. Notice the use of an “even though” locution in the following:

Even though it felt like I could either choose to donate to *Castanea Dentata* or choose to donate to *Ulmus Dentata*, if causal completeness is true then I couldn't really have chosen differently than I did. (Emphasis added.)

Although this seems to be a natural phrasing of the question, one might think that “even though” primes the participant to agree with the statement, which in this case is an incompatibilist response. The final study drops this potentially troublesome phrase and also tests whether the phenomenology of epistemic uncertainty differs from the phenomenology of being able to do otherwise in terms of compatibility with determinism. Condition 1 again presented participants with an actual choice among two options and tested whether they would continue to report having an incompatibilist phenomenology as of being able to do otherwise. Condition 2 focused on epistemic phenomenology.

Method

Participants:

106 participants were initially recruited online through the Mechanical Turk (MTurk) website. The survey itself was conducted using SurveyMonkey. 15 participants did not complete the survey, or indicated that they had recently taken a “very similar” survey. They were excluded from the analysis.

Materials:

The vignette and first question for condition 1 read as follows.

Part 1: The ability to do otherwise

In both conditions, participants were told that they would have a chance to win 5 cents if they picked the right button. The text went as follows:

At the bottom of this page, there are two buttons, labeled H and V. Each option is currently available for you to choose. In a moment, we'll ask you to choose just one of them. For this survey, only one of the buttons will give you an extra \$0.05 (as bonus payment on MTurk)

if you choose it. But we won't tell you which button it is—you'll have to make a choice and find out.

But don't decide just yet.

First, consider how things seem to you as you face your decision. In particular, consider what it's like to decide which option to choose.

In condition 1, participants were asked to indicate agreement (on a 1–7 scale) with the following statement:

When deciding which option to choose, it feels like I can either choose H or choose V.

Condition 2 was the same except that we dropped the modal “can” and asked participants to “consider what it's like to wonder which option you'll choose.” Participants were then asked to indicate their level of agreement with a statement about *epistemic* phenomenology:

When wondering which option I'll choose, it feels like I don't know for sure before I select a button which button is the bonus button.

As in study 1, only participants who agreed with the ability-to-do-otherwise statement proceeded to the training.

The two available options—H and V—appeared at the bottom of the screen, with a radio button representing each option. Participants were not told whether they had chosen the bonus button (H) until after they had answered the compatibility question.

Part 2: Training on determinism

The training section was the same as that used in study 1, and again participants only proceeded to the compatibility question if they passed the training.

Part 3: Consistency

The compatibility question was adjusted for the new cases. In condition 1, participants were told:

Now, recall the button-choosing situation. You previously agreed with the following statement:

When deciding which option to choose, it feels like I can either choose H or choose V.

Considering this previous statement about how things felt to you before your choice and your understanding of causal completeness, please indicate your level of agreement with the following:

If causal completeness is true, then I couldn't really have chosen differently than I did.

In condition 2, participants were reminded that they agreed with this statement:

When wondering which option I'll choose, it feels like I don't know for sure before I select a button which button is the bonus button.

They were then asked to indicate their level of agreement with the following:

If causal completeness is true, then I knew for sure before I selected a button which button was the bonus button.

The aim was to test whether participants distinguish the sort of alternative possibilities they reported themselves as experiencing in other conditions from clearly compatibilist alternative possibilities, which have to do simply with our ignorance of the future.

Results

Of the 53 participants who started condition 1, 47 completed it and had not recently taken a very similar survey (2 had). Of these, 44 indicated a phenomenology as of there being alternative possibilities in the situation and all but 3 of them passed the training section.¹⁹ The remaining 41 participants gave a mean response of 5.34 on the compatibility question, which differed significantly from the midpoint of the scale, $t(40) = 4.54$, $p < .001$. That is, participants once again demonstrated a strong tendency to interpret their agentic experience as being incompatible with determinism.

In condition 2, of the 53 participants who started the survey, 44 completed it and had not recently taken a very similar survey (8 had). Of these, 39 indicated a phenomenology of an ability to choose among possibilities and all but 1 of them passed the training section.²⁰ The remaining 38 participants gave a mean response of 2.66 on the compatibility question, which differed significantly from the midpoint of the scale, $t(37) = -5.23$, $p < .001$. That is, participants tended to regard their phenomenology of uncertainty about the future as compatible with determinism. A t -test between conditions 1 and 2 showed that results differed significantly between these two conditions, $t(76) = 6.85$, $p < .001$.

This final study provides yet further evidence that people do indeed judge that their experience of deciding is inconsistent with determinism, in the sense that the experience is non-veridical if determinism is true. At the same time, people tend to think that the feeling of not knowing what will happen is perfectly consistent with determinism. This suggests an appropriate sensitivity to the fact that ignorance is not incompatible with determinism.

¹⁹ Eight participants required correction, and passed the training section. There was a significant difference in responses between those who required correction and those who didn't. Those who required correction reported that their phenomenology was incompatible with CC ($M = 4.25$) but to a lesser degree than those who answered these questions correctly the first time ($M = 5.60$), $p = .057$.

²⁰ Nine participants required correction, and passed the training section. There were no statistically significant differences in responses between those who required extra training and those who didn't ($p = .2$).

2.5 General Discussion

2.5.1 Incompatibilism

These results have implications for several issues concerning free will. Perhaps most importantly, these studies seem to vindicate the incompatibilist descriptions of the experience of being able to do otherwise suggested by Campbell, O'Connor, and Searle. By the same token, the results run counter to the compatibilist descriptions of our experience suggested by Mill, Grünbaum, and Nahmias and colleagues. The design of these studies left it open for participants to describe their experience as involving the ability to do otherwise, while allowing them to interpret this ability however they wished. The results indicate that the people in the population that was tested tended to judge that their experience was *incompatible* with determinism.

The results also address a concern that has plagued recent work on intuitions about free will. Nahmias and Murray (2011) contend that people give incompatibilist responses in previous experiments simply because people misunderstand determinism. This is an important concern. But rather than merely testing to see whether people misunderstand determinism, my colleagues and I attacked the comprehension issue directly by exploiting the familiar technique of training to criterion. And we did not find any widespread confusion of determinism and bypassing. Part 1 of the training controls for confusion between determinism and fatalism. And the majority of participants reported that the accuracy of their experience as of being able to do otherwise is inconsistent with *determinism*, correctly understood. Across all the studies, the percentage of participants who didn't make it to the compatibility question due to failing the training section was small, at 6.15%. When we look at those participants who answered part 1 of the training incorrectly—that is, at those who *did* initially confuse determinism with fatalism, and who were directed to the follow-up training question—the percentage was small compared with Nahmias and Murray's results: Only 20.68% of participants initially made this mistake. Of those who initially made the mistake,

85.71% answered the follow-up training question correctly. Thus, fewer than 3% of participants continued to confuse determinism and fatalism after training. And those who required correction did not respond in any significant way differently from those who didn't.²¹

2.5.2 The Ability to Do Otherwise

Much of the free-will debate, since at least Hobbes, has been about an *ability to do otherwise*. One influential compatibilist thought is that the notion of the ability to do otherwise should be understood in contrast to constraint or coercion. The idea is that an agent is able to do otherwise just in case, if she had chosen, or wanted, or tried to do otherwise, then she would have done so (Cf. Moore 1912). There are also recent versions of such a “conditional analysis” of the ability to do otherwise. According to Kadri Vihvelin (2004), for instance, an ability to act (or not to act, which is simply to be able to act in another way) is analyzable along something like the following lines: an agent can ϕ at t_1 (say, raise her hand at t_1) just in case were she to choose to ϕ at t_2 , and her body stayed working normally and nothing interfered with her, she would ϕ at t_2 .²² In other words, Vihvelin holds that “persons have abilities by *having intrinsic properties that are the causal basis of the ability*” (2004: 438). So Vihvelin thinks that an ability to act is a disposition, or a bundle of dispositions. And, as she points out, “no one denies that dispositions are compatible with determinism” (2004: 429). After all, even if determinism is true, glass is still fragile—i.e., it has the disposition to break if struck.²³

²¹ Again, with the exception of study 2, condition 1, and study 3, condition 1. (See footnotes 16 and 19.)

²² Vihvelin's exact formulation is as follows: “S has the ability at time t to do X iff, for some intrinsic property or set of properties B that S has at t , for some time t' after t , if S chose (decided, intended, or tried) at t to do X, and S were to retain B until t' , S's choosing (deciding, intending, or trying) to do X and S's having of B would jointly be an S-complete cause of S's doing X” (2004: 438).

²³ For similar accounts, see Smith (2004) and Fara (2008). Questions persist (see e.g., Clarke 2009) about whether any “dispositionalist” account is an adequate analysis of the ability to act, and thus of the ability to act otherwise.

Other compatibilists embrace an epistemic reading of “can do otherwise.” On this view, to maintain that I can go left or right is simply to note that it is epistemically open whether I will go left or right. J. J. C. Smart argues that this is a natural way to interpret the expression “could have done otherwise” even outside the sphere of action. When I say, “the plate fell, and it could have broken,” I am not, says Smart, committing myself to any claim about determinism. Rather, what I am saying is that, before the plate completed its fall, for all I knew, the plate would break (1961: 298). Similarly, perhaps when I say that Oswald could have done otherwise, all I’m saying is that, before Oswald pulled the trigger, for all anyone knew, he wouldn’t pull the trigger. If I’m merely making a claim about epistemic possibilities, then there is no conflict with determinism.

By contrast, incompatibilists think being able to do otherwise (in the relevant contexts) means being able to do something other than what one does, all prior conditions (including one’s desires) remaining the same. This ability is presumed to be a matter of fact, not something about our epistemic access to facts.

At least insofar as the relevant notion of the ability to do otherwise is reflected in the experience as of being able to do otherwise, my results suggest that the compatibilist accounts fail. Across three studies, participants tended to interpret their agentic experience in terms of an ability to do otherwise, and they interpreted that ability incompatibilistically. Concerning the traditional compatibilist analysis, the results equally undercut old and new versions. After all, participants were allowed to describe their experiences as involving the ability to do otherwise or not, where they were free to interpret this ability however they wished. Participants then judged that *this* ability—the one they had been allowed to interpret however they wished—was incompatible with determinism. The epistemic compatibilist account is also undermined by these results. Participants gave compatibilist judgments about the case of ignorance about the future (study 3, condition 2), indicating that they do have an appreciation that the feeling of uncertainty is consistent with determinism.

It seems reasonable to conclude that the notion of “can do otherwise,” at least with respect to one’s decisions, is naturally interpreted in ways that contravene the most familiar compatibilist approaches in the philosophical literature. When participants attend to their experience while they consider future events, their usage of “can” tends to reflect a sense of *metaphysical* openness that is incompatible with determinism.

2.5.3 Misinterpreting One’s Agentive Experience

Obviously, the fact that people interpret their agentive experience as incompatibilist doesn’t show that people actually have an incompatibilist ability to do otherwise. Terry Horgan argues that people might be mistaken in their interpretation of their own phenomenology. He allows that people might regard their agentive experience as incompatibilist:

When one attends introspectively to one’s agentive phenomenology, with its... [representational]... aspects of freedom... and when one simultaneously asks reflectively whether the veridicality of this phenomenology is compatible with causal determinism..., one feels *some* tendency to judge that the answer to such compatibility questions is No. (Forthcoming)²⁴

But Horgan notes that we must distinguish between the content of our experience and the content of judgments. The former kind of content Horgan dubs “presentational content,” and it

...is the kind that accrues to phenomenology directly—apart from whether or not one has the capacity to articulate this content linguistically and understand what one is thus articulating, and apart from whether or not one has the kind of sophisticated conceptual repertoire

²⁴ For Horgan, this presentational “aspect of freedom” is what I have been calling the experience as of being able to do otherwise.

that would be required to understand such an articulation. (Forthcoming)

By contrast, “judgmental content” is the kind of content associated with linguistic articulations. Of course, we make judgments about our phenomenology, and so we can have judgmental content that aims to capture our presentational content. The key point here is that it is possible for our judgments about the (presentational) content of our experience to go awry.

That said, those judgments are at least *prima facie* evidence of the nature of the presentational-cum-phenomenal content, so one would need some positive reason to think that participants have systematically misinterpreted the nature of their phenomenology. Further, even if it is granted *arguendo* that the presentational content of agentive experience is (in the first instance) compatible with determinism, and that reports to the contrary count as mistaken interpretations, still, the fact that people judge the experience incompatibilist would be significant. For one thing, when considering how best to understand the notion of the “ability to do otherwise,” in many cases what will be of primary importance is how people *think* about their ability to do otherwise, and that is clearly judgmental. Second and more interestingly, judgmental content can feed back into presentational content. It is well known that what one judges about a situation can affect one’s perception of the situation. Horgan recognizes this, and he notes that the distinction between presentational and judgmental content isn’t always sharp: “it may very well be that the two kinds of content can interpenetrate to a substantial extent” (in press). As a result, even if the presentational content of agentive experience is, in the first instance, compatibilist, that doesn’t mean that the presentational content *remains* compatibilist. It might be that the incompatibilist judgment shapes the presentational content.²⁵

²⁵ Note that Horgan’s notion of “presentational content” is not simply “raw feels” with no propositional content. For everyone would concede that insofar as we have incompatibilist phenomenology, it must be presented at a level with greater conceptual sophistication than is provided by raw feels. Horgan is explicit about the possibility of rich conceptual resources being implicated in presentational content: “It is plausible. . . that humans can have presentational contents the pos-

2.5.4 Deliberation Compatibilism

A final issue that might be illuminated by these results is the debate over the presuppositions of deliberation. Some philosophers have maintained that deliberation carries with it a presumption of genuinely open possibilities of an incompatibilist variety. Richard Taylor writes, “I cannot deliberate about what to do, even though I may not know what I am going to do, unless I believe that it is up to me what I am going to do” (1983: 38–39). And this “up to me” is incompatible with determinism. Van Inwagen makes a similar point: “[I]f someone deliberates about whether to do A or to do B, it follows that his behavior manifests a belief that it is possible for him to do A—that he can do A, that he has it within his power to do A—and a belief that it is *possible* for him to do B” (1983: 155).

On the other side, we find “deliberation compatibilists,” who maintain that deliberation contains no such presuppositions. Tomis Kapitan begins his paper (which would become the locus classicus for deliberation compatibilism) thus:

By *deliberation* we understand practical reasoning with an end in view of choosing some course of action. Integral to it is the agent’s sense of alternative possibilities, that is, of two or more courses of action he presumes are open for him to undertake or not. (1986: 230)

Kapitan goes on to argue that the presumption of openness does not require *metaphysical* openness, but only *epistemic* openness.²⁶ A number of philosophers have followed Kapitan in developing compatibilist accounts of the presuppositions behind deliberation (e.g., Nelkin 2004, Pereboom 2008).

session of which require (at least causally) a fairly rich repertoire of background concepts that can figure in judgmental states.” For instance, “One can have presentational experiences, for instance, as-of computers, automobiles, airplanes, train stations” (Horgan forthcoming).

²⁶ According to Kapitan and other deliberation compatibilists, there are other conditions, too. In particular, Kapitan maintains that deliberation carries a presupposition of *efficacy*, which he characterizes roughly as follows: “an agent presumes that his ϕ -ing is an open alternative for him only if he presumes that he would ϕ if and only if he were to choose to ϕ ” (1986: 234). See also Pereboom (2008: 288). I leave this complication aside since it doesn’t affect my point.

Insofar as deliberation compatibilism claims that deliberation is not *as a matter of fact* experienced as having incompatibilist presuppositions, the studies reported here indicate that this position is mistaken. This does not decide the dispute concerning deliberation compatibilism, but it does show that one should distinguish three versions of deliberation-compatibilism:

- (1) people's beliefs about their current deliberations are compatible with determinism;
- (2) people's beliefs about their current deliberations are not compatible with determinism, but they can be adjusted to be compatible;
- (3) people's beliefs about their current deliberations are not, and cannot be adjusted to be, compatible with determinism, but we can conceive of a rational being whose beliefs about deliberation are compatible with determinism.

The results of the studies that my colleagues and I ran suggest that the first version of deliberation compatibilism is false. People's beliefs about their deliberations are incompatibilist. The second version—that our actual experiences are incompatibilist but revisable—is an interesting possibility, but it remains an open question whether it is possible to revise this aspect of our experience. Until we know more about what generates the incompatibilist experience, it is hard to know whether it can be modified. One possibility is that the incompatibilist experience is generated in a way that is not cognitively penetrable (see e.g., Bayne 2011). That is, it might be that even if we form the explicit high-level belief that deliberation is theoretically compatible with determinism, this will not eradicate our experience of our deliberation as incompatibilist. The third version of deliberation compatibilism—that we can conceive of rational creatures who deliberate as determinists—is not under any threat from our results. But if it turns out to be impossible for us to *be* such rational animals, that might undercut some of the interest of deliberation compatibilism.

2.6 Conclusion

The experience as of *being able to do otherwise* has long been central to debates about agency and free will. Libertarians appeal to this experience as evidence that determinism is false; compatibilists reject the libertarian accounts of the character of the experience. Despite the pivotal role of experience in these arguments, the experience itself has received scant attention. The studies reported here are an attempt to advance the issue. My colleagues and I found consistently incompatibilist judgments about the nature of the experience as of being able to do otherwise. This lends support to the phenomenological claim of libertarians, though I am not inclined to take the phenomenology of indeterminism as evidence that agency isn't determined (as will become clear in chapters 3–5). The results of these studies also suggest that existing compatibilist interpretations of the notion of “ability to do otherwise” are not adequate to people's reported experience as of being able to do otherwise. Finally, the results that I have reported here also speak to the presuppositions of deliberation. What my colleagues' and my studies indicate is that *as a matter of fact* our experience of deliberation features metaphysical openness (that is inconsistent with determinism). While this does not decide the dispute between deliberation compatibilists and deliberation incompatibilists, it does make salient the possibility that deliberation compatibilism requires an account of deliberation that is explicitly revisionist with respect to people's actual experience of deliberation.

As a result, the answer to the descriptive question about people's experience of modal freedom appears to be incompatibilist. We saw in Chapter 1 that the descriptive question is an empirical question about people's pre-theoretic belief-tendencies about their own and others' agency. We also saw that there are at least two ways to address this question: (1) ask people about their intuitions or beliefs regarding modal freedom, or (2) ask people about their *experience* of being modally free. In this chapter, I have focused on the second method, concerning experiences of freedom. As a matter of fact, people actually tend to describe their experience of modal freedom (i.e., of being able to do otherwise) as inconsistent

with determinism, in the sense of being non-veridical if determinism is true. On the further plausible assumption that experiences of modal freedom partly drive people's beliefs (or intuitions) about freedom, these results provide support for the view that people's pre-theoretic belief-tendencies are at least partly incompatibilist.

In the next chapter, I defend these empirical results against an error theory for incompatibilist judgments about experience developed by Terry Horgan. I argue that compatibilists should simply grant that the answer to the descriptive question about experiences of modal freedom is incompatibilist. Even so, I propose a prescriptive compatibilist story about the veridicality of experiences of modal freedom. I argue that despite granting (at least *arguendo*) that our experience of modal freedom has libertarian content, our experience might still be veridical under the assumption of determinism. I sketch this view at the end of Chapter 3 and develop it more fully in Chapter 4. Finally, in Chapter 5, I develop a plausible (but deflationary) etiological story about the source of libertarian experiences and beliefs.

Chapter 3

Is Agentive Experience Compatible with Determinism?

Many philosophers think not only that we are free to act otherwise than we do, but also that we *experience* being free in this way. Terry Horgan argues that such experience is compatibilist: it is veridical even if determinism is true. According to Horgan, when people judge their experience as incompatibilist, they misinterpret it. However, while Horgan's position is attractive, it incurs significant theoretical costs. I sketch an alternative way to be a compatibilist about experiences of free agency that avoids these costs while also exhibiting considerable advantages of its own.

3.1 Introduction

The waiter offers you ice-cream. "Chocolate or vanilla?" he asks. Each flavor is delicious, but you know you should only choose one. You hesitate. It feels like you are free to choose vanilla. Nevertheless, it also feels like you can refrain from choosing it—say, by choosing chocolate instead. It feels like you are free to do otherwise.

Is this experience accurate, or veridical, assuming determinism? Many in-

compatibilists, who think the freedom to do otherwise is incompatible with determinism, have thought it is not.¹ They think that we experience having a freedom that is inconsistent with determinism. According to John Searle, for instance, our experience amounts to the feeling that “we could be doing something else right here and now, that is, all other conditions remaining the same” (1984: 95). Some incompatibilists—the libertarians—even go so far as to maintain that our experience in this regard is evidence that we *possess* an incompatibilist freedom (e.g., O’Connor 1995).

Compatibilists think that the freedom to do otherwise is consistent with determinism (Moore 1912; Vihvelin 2004; Fara 2008). If the freedom we experience possessing is compatibilist, then our experience of being free to do otherwise is accurate, assuming determinism. For instance, compatibilists often suggest that we experience freedom conditionally: in the above example, as long as we are free from constraint, coercion, and an addiction to vanilla ice-cream (say), our experience is that we are free to choose chocolate *if* we want (or try) to do so, and similarly regarding vanilla (Cf. Mill 1865; Grünbaum 1952; Nahmias et al. 2004). If that is right, then it undermines a key motivation for libertarianism—the view that being free to do otherwise is inconsistent with determinism, and we have such freedom. After all, if the nature of our experience is compatibilist, then libertarians cannot argue from the incompatibilist nature of experience to our possessing an incompatibilist freedom.

A somewhat different compatibilist strategy is to grant that introspection *seems* to reveal that experience is incompatibilist, yet insist that introspection is not reliable in this domain. Terry Horgan adopts this strategy (forthcoming, 2012,

¹ Traditionally, the sort of freedom required for moral responsibility was thought to be the freedom or ability to do otherwise. Many philosophers—both compatibilist and incompatibilist—have now abandoned this idea. They think that the sort of freedom or control that is required for responsibility is that one be the appropriate *source* of one’s action, where this does not require being free to do otherwise. As I outlined in Chapter 1, call the freedom condition on responsibility *moral freedom*, and the ability to do otherwise *modal freedom*. Even if moral freedom does not require modal freedom—and some compatibilists continue to think it does (e.g., Vihvelin 2004; Fara 2008)—the question whether modal freedom is compatible with determinism presses on us in its own right.

2011, 2007). Horgan agrees that people often *think* that their experience is incompatibilist. However, he argues that even when people judge their experience as incompatibilist, actually it is compatibilist: people misinterpret their experience. By spelling out how this happens, Horgan provides an error theory for incompatibilist judgments about experience.

Horgan's compatibilist strategy is attractive and has important theoretical advantages. Yet it also incurs theoretical costs. After considering Horgan's position, I sketch an alternative way to be a compatibilist about the experience of freedom that avoids incurring these costs. On my view, even if we take people's incompatibilist reports about their experience at face value, and thus grant that such experience has genuinely incompatibilist content, there is still an important respect in which the experience is accurate, assuming determinism. Thus, I grant (at least for the sake of argument) an incompatibilist answer to the descriptive question that I outlined in Chapter 1, while defending compatibilism on the prescriptive question. As I explain in the next section (and in Chapter 4), the strategy I adopt does not require *defending* the idea that people's experience of modal freedom is incompatibilist, as is suggested by the results of the studies reported in Chapter 2. Instead, I argue that it turns out to be dialectically advantageous for compatibilists to grant that this is the case, at least *arguendo*. Thus, commitment to incompatibilist experience is simply taken hypothetically, for the sake of argument, in the present chapter and in Chapter 4. (In Chapter 5, I again take the existence of incompatibilist experience seriously, and I there develop a deflationary etiology for such experience.)

3.2 Agentive Experience

To forestall any confusion, let me begin by clarifying some terminology. Granting that we actually experience being free to do otherwise, experience-incompatibilists think that this experience is inaccurate if determinism is true. Call such experiences libertarian or incompatibilist. Experience-compatibilists

think the opposite: the experience might be accurate, even assuming determinism. Call such experiences compatibilist. Finally, call the question whether our experience of being free to do otherwise is compatible with determinism (in the way just outlined) the experience-compatibilism question.

I take an experience to be any representational mental state with phenomenal character, where phenomenal character is what-it's-like (or what it feels like) to be in that mental state. An experience's phenomenology is just its phenomenal character. The satisfaction conditions for an experience are its accuracy or veridicality conditions. For any experience, its content yields a veridicality condition: the content specifies how the world must be in order for the experience to be veridical. If a visual experience has the content *squareness*, where this property is attributed to a particular object, then that experience is veridical only if the object in question *is* square. Moreover, I assume a close tie between content and phenomenology, so that an experience's phenomenology shares a veridicality condition with its content. Call this phenomenal content. A visual experience of seeing a red apple will have the phenomenal character *reddishness*, and thus the content that a certain object one sees—the apple—is red. Such content is constitutively determined by the phenomenal character,² and it is veridical only if the apple actually is red.

The property of being able to act otherwise is the property of being free in a particular way: it is to possess a specific—not just a general—ability or freedom, as I explained in Chapter 1. To recap briefly: General abilities are uncontroversially compatibilist. I might possess the general ability to raise my hand an hour from now, without having the specific ability to exercise it just then, perhaps because I will be asleep. Determinism is compatible with my retaining these sorts of unexercised abilities. Only specific abilities are at issue in the question whether freedom is compatible with determinism. In Chapter 1, I characterized this notion as follows. One has the specific ability to do something only if (i) one has a

² I remain neutral about whether representational properties determine phenomenal properties, or vice versa.

general ability to do it, (ii) one has an opportunity to do it, and (iii) holding fixed one's motivations at the time (including the exact strengths of one's motivations), one can exercise one's general ability to act in that way at that time. If this is true of more than one available option at a time, then one has the specific ability to do otherwise.

What about agentic experience? For a start, one might wonder whether there *is* such a phenomenon. Even if there is, it is a further question whether high-level properties like *being able to do otherwise* feature in it. One might think that even if people believe that they are free to do otherwise, they do not *experience* such freedom. Yet, as we shall see, appeals made by philosophers to experience imply that such properties do feature in agentic experience. I will deal with these two issues in turn.

First, certain disorders of agency speak against the suggestion that agentic experiences in general are reducible to beliefs about agency (Cf. Bayne 2008: 185–87; 2011: 360). A patient with anarchic hand syndrome may discover her hand “doing” things—perhaps against her wishes. She may even describe her hand as “having a will of its own.” It is plausible that such a patient's disavowal of her hand's “actions” may be due to the fact that she does not experience its movements as issuing from her own agency. She might come to *believe* that these movements are, in a sense, her own actions (who else's could they be?). Yet this judgment will hardly affect her experience. If that is right, then agentic experiences are not beliefs about agency.³

Even granting agentic experiences, it is a further question whether high-level properties like *being able to do otherwise* feature in them. On a conservative view, the contents of agentic experiences contrast with those of other perceptual modalities, like vision (Cf. Bayne 2008: 189–90). On this view, agentic expe-

³ Bayne claims that such cases show that agentic experiences are not influenced by cognitive states (2011: 360). However, just because my experience of seeing a stick in the water seeming to be bent is not altered by my explicit belief that it is straight does not show that my visual experience is not shaped by my *background* beliefs. Likewise, I propose, for agentic experiences. I discuss this issue further in section 3.4.

periences merely represent one as acting, controlling one's action, and acting with effort (Cf. Horgan forthcoming; Bayne 2008: 190–92). By contrast, a more liberal view takes the contents of agentic experiences to be relatively rich. For example, they additionally represent one as being the source of one's actions, and one's actions as being free—including in the sense that one is free to do otherwise.⁴

At any rate, the way in which many people describe their agentic experience indicates that they experience being free to do otherwise, and such beliefs drive a good deal of philosophical theorizing about free will. Indeed, we have already seen that many philosophers take it as a basic insight that being free to do otherwise is central to the notion of human freedom, and our possessing such freedom is confirmed by our phenomenology of deliberation. This goes for both compatibilists and incompatibilists, as noted at the opening section of this chapter (as well as in chapters 1 and 2).⁵

Finally, the results of the experiments that I reported in Chapter 2 indicate that it is not just philosophers who report experiencing being free to do otherwise—ordinary people report such experiences as well. In these experiments, every effort was made to ensure that participants attended to their relevant experience (if any). Over 91% of participants reported experiencing being free to choose either way, where it was left open for them to interpret such freedom however they wished. These results support the claim that participants had an experience of being free to do otherwise.

Furthermore, participants in these studies went on to judge their experience as incompatibilist when the notion of determinism was explained to them. Following a series of comprehension checks, participants were asked whether the experience they had earlier reported was consistent with determinism. Across a range of conditions, participants said their experience was inconsistent. Whether

⁴ See Bayne (2008) for discussion of this issue. For analogous debates about whether high-level properties are represented in perceptual experience, see Siegel (2006) and Bayne (2009).

⁵ In addition to the authors mentioned in section 3.1, many other philosophers grant that there are identifiable experiences of being free to do otherwise. See for instance Hume (1960/1739), Reid (1788), C. A. Campbell (1951), Lehrer (1960), Ginet (1997), Nahmias et al. (2004), and Holton (2006).

participants merely imagined making a choice or actually made a choice, whether the choice was morally salient or not, or whether the choice was present-focused or retrospective, participants reported an experience of being free to do otherwise that is incompatible with determinism. Thus, apparently, people not only experience being free to do otherwise, but they also report that their experience in this regard is inaccurate if determinism is true. This offers at least prima facie evidence in favor of experience-incompatibilism, evidence that must be addressed by compatibilists.⁶

For compatibilists, such results are problematic for at least two reasons. First, if a compatibilist theory of freedom fails to capture our experience of being free to do otherwise, then it is not clear whether the theory is explaining the right phenomenon. It seems plausible that the freedom the theory explains should be the one that we experience possessing. Second, if all our experiences of being free to do otherwise are illusory, assuming determinism, then this leaves us with no way of distinguishing—*just in terms of the accuracy conditions of our phenomenology*—illusory experiences from experiences we normally think are accurate. Imagine that you wake at night and consider switching on the light. As you lie there, you experience being free to switch on the light, or to refrain from doing so—for instance, by intentionally remaining motionless. Surely, we want to distinguish this case from one in which, unbeknownst to you, you have been paralyzed by a drug, yet still you experience being free to switch on the light or to refrain from doing so. If all your experiences of being free to do otherwise are illusory *because* determinism is true, then there is no compatibilist way to make sense of the idea that your experience in the first case is accurate in exactly the way it is not in the second. Granting that we experience being incompatibilistically free, we are left with the verdict that the content of our phenomenology in

⁶ As we shall see, the claim that these participants' experiences were incompatibilist is tenuous in the present context, since Horgan's claim is precisely that people often misinterpret their experiences of freedom. By contrast, I will argue that even if these participants' experiences are incompatibilist in a certain respect, they might nevertheless be accurate in another respect, assuming determinism.

each case makes any such experience utterly illusory, assuming determinism. One might avoid this verdict if the experience had compatibilist content—for instance, if it were best described as an experience of conditional freedom. Yet, as we have seen, this appears to go against the evidence. Moreover, for the reasons given a moment ago, compatibilists cannot insist that the experience is theoretically unimportant. Compatibilists need an alternative strategy.

Terry Horgan proposes just such a strategy. Horgan agrees that the relevant experience is theoretically important, and he agrees that many people have a strong tendency to think that their experience is incompatibilist. However, he argues that even when people judge their experience as incompatibilist, actually it is compatibilist: people tend to misinterpret their actual experience. By spelling out how this happens, Horgan provides an error theory for incompatibilist judgments about experience. If Horgan is right, then the participant responses in Chapter 1 may well be in error.

Horgan's compatibilist strategy is attractive, and it has clear theoretical advantages. However, it also incurs significant theoretical costs. In section 3.3, I sketch Horgan's view, then in sections 3.4 and 3.5, I outline the costs that it incurs. After considering Horgan's position, I sketch an alternative way to be a compatibilist about the experience of freedom, which avoids these costs. My position also has distinct advantages over Horgan's. For instance, I argue that even if we take people's incompatibilist reports about their experience of freedom at face value, and thereby grant both that *introspection reliably latches onto the content of such experience* and that *such content is rich enough to be incompatibilist*, there is still an important respect in which the experience is compatibilist: it is veridical even if determinism is true.

3.3 Horgan's View

How does Horgan defend experience-compatibilism? According to Horgan, people's incompatibilist reports about their experience of freedom are in error. Now,

incompatibilist experience might, of course, be mistaken in the sense of being non-veridical. However, Horgan suggests another way in which people might be mistaken—they might misinterpret the relevant experience. Horgan concedes that people often *judge* their experience of being free to do otherwise as incompatibilist. For instance, he says,

When one attends introspectively to one’s agentic phenomenology, with its ... [representational] ... aspects of freedom ... and when one simultaneously asks reflectively whether the veridicality of this phenomenology is compatible with causal determinism ..., one feels some tendency to judge that the answer to such compatibility questions is No. (Forthcoming)⁷

Horgan thinks that while introspection is reliable in some domains, introspective judgments about whether our agentic experience is compatible with determinism are uniquely unreliable (see e.g., 2012; 2011). He begins by distinguishing between two sorts of introspection: (1) *attentive introspection*, which involves “paying attention to certain aspects of one’s current experience”; and (2) *judgmental introspection*, which involves “forming a judgment about the nature of one’s current experience” (2011: 84). The kind of content on which we attentively introspect is “presentational content,” which is “... the kind that accrues to phenomenology directly—apart from whether or not one has the capacity to articulate this content linguistically and understand what one is thus articulating, and apart from whether or not one has the kind of sophisticated conceptual repertoire that would be required to understand such an articulation” (forthcoming). This is the sort of content about which we make judgments when we judgmentally introspect.⁸

⁷ Horgan deals with several types of agentic experience. For the sake of simplicity, I focus solely on Horgan’s treatment of what he calls the “freedom” aspect of experience, which he understands as the experience of being free to do otherwise.

⁸ Presentational content is roughly what I am calling phenomenal content. As we will see, though, it is not *equivalent* to phenomenal content, since there may be sorts of phenomenal content

In judgmental introspection, we attend to certain aspects of our experience, and then form a judgment about them. Thus, “Judgmental introspection . . . deploys attentive introspection, while also generating a judgment about what is being attended to” (2011: 84). There is no appearance/reality gap when we attentively introspect. Yet, in judgmentally introspecting on our experience, we can, as it turns out, go wrong: we may be subject to what Horgan calls a labeling fallacy (2012: 408–409). For instance, we might make a performance error in applying the ordinary judgmental concept ‘red’ to our experience of redness: we might mistakenly apply the ordinary concept ‘green.’ In Horgan’s parlance, we might “mislabel” the phenomenology. Presumably, this hardly (if) ever happens. So, while attentive introspection is infallible, judgmental introspection about color experiences is not quite infallible, although it nearly is.⁹

The point for now is that Horgan thinks judgments about our agentic experience are not like this. It is not merely that our judgments are fallible; *they are not reliable at all*—at least when it comes to our experience of freedom and whether it is compatible with determinism. For a start, answering this compatibility question far outstrips what attentive introspection is capable of. The question can only be answered by means of judgmental introspection. Yet this process is uniquely ill-qualified for the task. Horgan thinks that when we try to answer the compatibility question about our experience of being free to do otherwise by judgmentally introspecting, we find that such introspection is unable to provide a reliable answer, even though the question is about the character of our introspectively available experience of freedom.¹⁰ It is not simply that we are subject to the occasional la-

other than what is “presented” in phenomenology. Even so, what Horgan calls “presentational content” is an important aspect (or sort) of phenomenal content. I leave aside this wrinkle since it does not affect what I have to say.

⁹ Horgan (2012; Horgan and Kriegel 2008) argues that there are also cases in which we are immune even from labeling fallacies when we make judgments about our experience—e.g., when we judge that “this experience has this feature.” Such judgments are infallible. These cases do not concern me here.

¹⁰ This is despite the fact that, for Horgan, experiences of freedom have intrinsic, determinate satisfaction conditions that are compatibilist. However, such compatibility is a “non-manifest”

being fallacy. Rather, we cannot reliably tell what the answer to the compatibility question *is* just by judgmentally introspecting.

Horgan thinks there are good reasons for this, which I will outline in a moment. Nevertheless, he admits that there are “sophisticated philosophers” who think that there is what he calls a “read-offable” incompatibilist answer to the compatibility question about determinism and the experience of being free to do otherwise, “since they have said so to me in philosophical discussion. And I confess to experiencing some temptation to think so myself, as I suspect you the reader do too—a temptation that needs explaining” (2012: 416). Horgan’s explanation has two parts.

For instance, if we are asked whether Alice is free to do otherwise than steal a piece of candy—when she is in possession of her faculties, is not being ordered to steal at gunpoint, is not subject to an irresistible addiction, and so on—we are liable to answer, “Yes.” By contrast, if we are asked whether Fred is free to do otherwise than lie in bed all morning, when—as it turns out—he is securely strapped down and is prevented from moving, we are liable to answer, “No.” By contrast with such cases, when it comes to knowing whether to apply the notion of being free to do otherwise to our experience while assuming the truth of determinism, we go beyond the limits our competence. Here, we are reasoning about whether the satisfaction conditions of our phenomenology are met under the assumption of a general hypothesis about the world, and we are liable to make mistakes. Instead, we should proceed by inference to the best explanation: seek a hypothesis that yields an answer to the compatibility question, while also explaining various “data,” such as the fact that we are normally competent in applying the notion of being free to do otherwise to concrete cases and without ever considering the thesis of determinism. Horgan thinks it reasonable to assume that such intuitive judgments result from our competence in applying the relevant notion, so that such judgments are “normally true” (forthcoming).

feature of the experiences. (Horgan allows that such experiences may have wide satisfaction conditions as well.)

Compatibilism accommodates these judgments easily, by enabling them to come out true even when we assume determinism. Incompatibilism, by contrast, requires that a more stringent condition be met—namely, that indeterminism (at a minimum) be true. So, Horgan thinks we should prefer compatibilism to incompatibilism. As he puts it, other things being equal “one hypothesis is better than another if it accommodates the attributional practices of competent users of the relevant concept better than the other” (forthcoming).

This is meant to show that any answer to the question whether our experience of freedom is compatibilist or not will go beyond the capabilities of judgmental introspection, and thus any such answer will be unreliable. Of course, Horgan also needs to explain why we have any tendency in the first place to judge our experience as incompatibilist. It is one thing to open up scope for error in our judgments, as Horgan does, but it is another to explain away this judgmental tendency. To this end, Horgan offers a two-part debunking explanation of incompatibilist judgments about experience. First, Horgan suggests a way in which we might introspectively confabulate. Second, he tells a contextualist story about the application conditions of the notion of freedom, which also applies to judgments about our experience of freedom. I will take these in turn.

First, Horgan suggests that if we think we can tell by introspection that our experience is incompatible with determinism, this may reflect a form of “introspective confabulation.” After all, it is one thing for me to know (A) by introspection:

(A) My experience does *not* present my behavior as determined by my prior states.

Yet it is another thing for me to know (B) simply by introspecting on my phenomenology:

(B) My experience presents my behavior as *not* determined by my prior states.

Horgan thinks that we can ascertain whether (A) is true by introspecting. However, (B) is distinct from (A), and we cannot ascertain whether (B) is true by introspection. Even if (B) were true, we could not know this merely by judgmentally

introspecting on experience. When we judge that our experience is incompatible, and thereby assert (B), either we are mistakenly inferring (B) from (A), or conflating (A) and (B).¹¹

Second, Horgan has a contextualist story to tell about the application conditions of the notion of freedom, which also applies to judgments about our experience of freedom:

I maintain that many concepts that figure importantly in philosophical problems are governed by implicit, contextually variable, semantic parameters—and that some forms of philosophical puzzlement arise largely because (i) posing a philosophical problem can tend to shift the implicit parameters toward settings under which the claims made using a given concept are more “demanding” in their truth conditions than the claims that would normally be made using that concept, and (ii) one tends not to notice this shift of the “score in the language game” when one is contemplating the philosophical problem. ... I maintain that the very posing of the question whether human freedom is compatible with ... determinism tends to alter the contextually operative settings on certain implicit semantic parameters that govern the concept *freedom*—and tends to drive those parameter settings so high that, in the newly created context, no item of behavior that is ... determined counts as *free*. (2007: 21–22)

These contextual parameters do not apply to phenomenal content. After all, it is plausible that non-human animals also have agential phenomenology, despite their mental content not being governed by contextual semantic parameters. Even so, Horgan thinks that when we introspect on our experience of freedom, while asking ourselves whether its content is compatible with determinism, our judgment gets “infected” by scorekeeping confusions, just as happens when we ask

¹¹ Alternatively, we might just be over-interpreting the character of what we introspectively attend to.

compatibility questions about the judgmental content of ordinary notions like ‘freedom.’

3.4 Problems with Horgan’s View

Even if we grant Horgan’s hypothesis about introspective confabulation, more needs to be said about how this sort of mistake occurs. For example, Shaun Nichols notes that generally people do not make this sort of mistake when it comes to headaches: “the phenomenology of headaches doesn’t present us with a set of deterministic headache-causes, but we don’t leap to indeterminist conclusions there” (2012: 296). In other words, we do not mistakenly infer from the claim that our experience does *not* present our headache as determined the further claim that we experience our headache as *not* determined. Thus, Horgan needs to say how the phenomenology of deliberation is relevantly different from that of headaches. This requirement is a theoretical cost of Horgan’s view, which the alternative position that I outline in section 3.6 (and develop further in Chapter 4) does not incur, since my view does not claim that incompatibilist judgments about experience are mistaken. Of course, I need to say something about the source of incompatibilist experiences, whereas Horgan does not. Yet my position has resources in this regard, as I outline later in this section.

Note too that Horgan’s proposal is backward-looking, since it focuses on our introspective access to the causes of our decisions. By contrast, deliberation and action-planning are importantly forward-looking. When we face decisions, our primary interest as deliberating agents is not the causal antecedents of our decision, but rather the alternatives with which we are presented, and our sense of being free to decide between them. These are some of the aspects of agentive experience that libertarians most often cite as incompatibilist, and they are also of concern to compatibilists. When we focus on the future, the content of our experience is not that our behavior is not determined by our prior states and the laws of nature, since for one thing our experience does not concern the laws of nature. More plausibly, our experience has a content, P, that is in fact incompatibilist, where P

is something like *openness to the future*. Notably, even some compatibilists grant that our experience of such openness is non-veridical if determinism is true. For instance, Keith Lehrer has claimed that when it comes to the way in which the future feels open, the incompatibilist “accurately describes what I find by introspecting, and I cannot believe that others do not find the same” (1960: 150). Such a paradigmatic compatibilist as David Hume (1960/1739) agrees with this sentiment when he writes, “There is a false . . . experience . . . of the liberty of indifference” (Bk. II, Part III, §II). This sense of openness to the future has been characterized by the (semi-)compatibilist John Fischer as like a “Garden of Forking Paths.” This metaphor portrays our options as arrayed before us like branches in a path, each one seemingly a realizable extension of the actual present into the future (Fischer 1994: 190). Fischer thinks that being free to do otherwise is *incompatible* with determinism, since for him determinism has the consequence that there is only one possible extension of our actual present into the future. For Horgan, determinism does not have this consequence. Although Horgan thinks that experiences of freedom are aptly described by metaphors like “Garden of Forking Paths,” and he agrees that libertarian descriptions of agentive experience are phenomenologically apt, he still thinks that it remains open whether the satisfaction conditions of such experience are compatibilist. However, nothing in his position speaks to this view of what the content of our agentive experience amounts to. This is another theoretical cost of his view, since there is at least *prima facie* reason to think that what it amounts to entails indeterminism. By contrast, my position avoids this cost by granting (at least for argument’s sake) that our experience of freedom has libertarian content, yet is normally veridical even if determinism is true.

There is also a further worry for Horgan’s position: judgments about experience can feed back into the experiences themselves. It is widely known that what one judges about a situation can affect one’s perception of the situation, via cognitive penetration. Roughly, cognitive penetration occurs when the phenomenal character of one’s experience is altered by one’s cognitive states—for instance, by one’s background beliefs. There is ample evidence that this actually occurs

in cases of visual perception.¹² Something similar may occur in agentic experience. For instance, one's background beliefs may influence the character of one's experience of freedom. Indeed, Horgan appears to recognize this possibility, and he notes that the distinction between phenomenal and judgmental content is not always sharp: "it may well be that the two kinds of content can interpenetrate to a substantial extent" (forthcoming). Thus, even if one's phenomenal content is initially compatibilist (as Horgan claims it is), that does not mean that it *remains* compatibilist. The incompatibilist judgment might shape the "presentational" content.

Whether this concern has bite depends on whether cognitive penetration actually occurs in agentic experience. Although there is no definitive evidence of this, certain considerations suggest that it is not just possible, but likely. First, to say that experience is cognitively penetrable is just to say that it is theory-laden, which is widely accepted.¹³ Indeed, it would be somewhat surprising if agentic experience turned out to be exceptional in *not* being theory-laden, or subject to top-down processing.¹⁴ Second, there is evidence that beliefs about determinism and free will have measurable effects on both cheating and punishing behavior in experimental settings. Priming participants to believe that determinism is true appears to result in their cheating more (Vohs and Schooler 2008), whereas priming them instead to believe that neural mechanism is true results in their punishing others less than when they retain the belief that people have whatever sort of freedom is (for them) undermined by neural mechanism—presumably, libertarian free will (Shariff et al. forthcoming). Such evidence, taken together with the fact that experience often tends to be theory-laden, suggests that background beliefs—for

¹² See e.g., Delk and Fillenbaum (1965) and Levin and Banaji (2006) for classic experiments that support the thesis of cognitive penetration. See also MacPherson (2012) for recent discussion of this phenomenon.

¹³ See e.g., Hanson (1958) for an early discussion of the theory-ladenness of perception and its bearing on science.

¹⁴ See Palmer (1999, Ch. 9) for a detailed discussion of top-down processing in everyday perceptual experience. The claim here is that a good deal of early-stage perceptual information is subject to subsequent top-down processing, whereby experience is influenced by concepts, beliefs, and expectations.

instance, about the falsity of determinism or the obtaining of an indeterminist or libertarian freedom—might well influence people’s agentive *experiences* of freedom.

Of course, if an explicit incompatibility judgment is what affects our experience, then one might object that before we make such a judgment our experience will not have been altered in this way. Yet this objection is mistaken. Let us grant that experiences of being free to do otherwise have a content, P, that is in fact incompatibilist, and which results from cognitive penetration. Even so, such content does not have to be caused by any *explicit* belief. It might instead be caused by a *background* assumption implicitly held about P. This is the standard view of what happens in visual experience (Cf. Delk and Fillenbaum 1965).

Conversely, someone might think that if, so to speak, the judgmental tail is wagging the experiential dog, then that is far less worrisome to compatibilists than if the relevant experience were inherently incompatibilist. Compatibilists can simply grant that these incorrect judgments result in incorrect experience, yet insist that the experience can be altered to be veridical once the relevant judgments are corrected. However, it is not clear that such correction will always (or even ever) be possible. After all, my visual experience of a stick in the water seeming to be bent is not altered by my forming the explicit belief that it is straight. Likewise, as we saw in section 3.2, the fact that a patient with anarchic hand syndrome forms the explicit belief that her hand’s movements are, in a sense, her own actions does not affect her alienated experience of these actions not being her own (Cf. Bayne 2008: 185–87; 2011: 360).

Finally, one might think that even if experiences of being free to do otherwise are theory-laden, and as a result many people end up with (perhaps unalterably) incompatibilist experience, still people’s experience beforehand was compatibilist. This may be so. Yet agents for whom this is the case will still have agentive experience that is genuinely non-veridical if determinism is true. To grant that much, for the experience-compatibilist, seems to give the game away.

However, it bears repeating that this worry has bite only if it turns out that agentic experience is in fact theory laden. Yet given the relative likelihood of this being the case, it would be theoretically advantageous for a compatibilist view about agentic experience to be immune to the worry. Horgan's position is not immune. By contrast, the alternative compatibilist position that I sketch in section 3.6 (and develop more fully in Chapter 4) *is* immune. Indeed, the possibility that agentic experience is theory laden serves to bolster my view, by positing a possible explanation for the source of people's incompatibilist experiences of being free to do otherwise.

3.5 Problems with Horgan's Contextualism

There are also problems with Horgan's contextualist proposal, the most important of which is that global worries about whether we are free—even in the sense of being able to do otherwise—can arise when contextual parameters are normal. If that is right, then Horgan's claim that such worries arise only when we raise the parameters beyond their normal settings and explicitly ask the compatibility question is false. Rather, it seems that our competence in applying the relevant notion of freedom is such that it might enable us to answer the compatibility question, and to do so reliably. That, in turn, would undermine Horgan's claim that incompatibilist answers to the compatibility question are the result of a scorekeeping confusion

On the sort of contextualism that Horgan seems to adopt, 'can' (or 'is able to') may be used with varying degrees of stringency. According to David Lewis (1986):

To say that something can happen means that its happening is compossible with certain facts. Which facts? That is determined, but sometimes not determined well enough, by context. An ape can't speak a human language—say, Finnish—but I can. Facts about the anatomy and operation of the ape's larynx and nervous system are

not compossible with his speaking Finnish. But don't take me along to Helsinki as your interpreter: I can't speak Finnish. My speaking Finnish is compossible with the facts considered so far, but not with further facts about my lack of training. What I can do, relevant to one set of facts, I cannot do, relative to another, more inclusive set. Whenever the context leaves it open which facts are to count as relevant, it is possible to equivocate about whether I can speak Finnish. (77)

Lewis's view is *contextualist* since the meaning of 'can' does not, by itself, determine which facts are relevant; the additionally relevant facts are determined by context. While all uses of 'can' share a common semantic element—they express compatibility with certain facts—the precise meaning of a particular use depends on something else, which Lewis calls "context," and which we can take to be *whatever facts determine the precise meaning of a particular use of 'can.'* So, 'S can A' means that S's A-ing is compatible with certain facts, where the relevant facts depend on the stringency with which 'can' is used. The sense in which determinism makes it impossible for someone to do anything other than what she does is this: given the actual history and laws, it is not physically possible for her to do anything else. Obviously, a similar contextualist line can be run for claims about being able to do otherwise.

Horgan maintains that we go beyond the limits of our competence when it comes to applying notions like *being able to do otherwise*, at least while assuming determinism. Here, we are reasoning about whether the application conditions of this notion are met under the assumption of a general metaphysical hypothesis, and we are liable to make mistakes. We do best, Horgan thinks, to proceed instead by inference to the best explanation; and that, he thinks, favors compatibilism.

Still, global worries about whether we are free—even in the sense of being able to do otherwise—can arise even when contextual parameters are *not* limit-case, as they are when we explicitly ask the compatibility question. Such worries can arise even when contextual parameters are normal, as they are when we ask

whether someone is free while making only whatever assumptions are appropriate to everyday situations.

Consider, for instance, the sort of case made famous by Harry Frankfurt (1969). In Frankfurt's original case, Black—a neurosurgeon—wants Jones to choose A. Black can intervene to control Jones's brain processes should Jones be about to choose B. Yet Black prefers not to intervene unnecessarily. Instead, he waits to see how Jones will choose on his own. Jones is unaware of Black's presence. Frankfurt claims that Jones lacks alternative possibilities in the case. However, if Jones chooses A on his own, he is apparently responsible for his choice even though he has no alternative, given that Black is ready to intervene to control Jones's brain processes.

Now assume that Jones's choosing A is choosing to kill Smith. A natural response to this case is to ask whether it is reasonable to expect that Jones have done something else instead, given that the conditions in which he found himself ruled out any alternative—and this despite the fact that these conditions play no role in why Jones *does* kill Smith (Cf. Widerker 2006). If we think it reasonable to expect that Jones not have killed Smith, then it seems we have located a conflict in our thinking about how to apply the notion of freedom. On one hand, if we consider the case just by focusing on the intervener, without ever considering determinism, we might want to grant—given that Black did not intervene—that Jones freely killed Smith. After all, he killed Smith on his own. On the other hand, it is not clear whether it is reasonable to expect that Jones have done something else instead. Recall, he was unable to do otherwise. Did Jones freely kill Smith? Perhaps we do not know. Have we illicitly raised the contextual parameters governing application of the relevant notion of freedom? It is not clear that we have. Once we point out that determinism is meant to function in the same way as Black, by blocking the availability of alternatives and thus blocking the ability to do otherwise, we have generated a global worry according to the ordinary standards governing application of the notion of freedom.¹⁵ The standards are

¹⁵ This is the standard way to understand how Frankfurt-cases work. Nevertheless, one might

ordinary since they do not invoke determinism, yet if Black functions in the same way as determinism (see footnote 15) then all we have to do in order to generate a global worry is to ask someone to imagine that there is *always* a figure like Black lurking in the background whenever anyone deliberates about doing anything.

If global worries about whether we are free arise from our competence in applying the relevant notion of freedom, even when the parameter settings are normal, then Horgan's contextualist move proves doubtful.¹⁶ Even when we apply ordinary standards, which are internal to our everyday judgmental practices and do not explicitly invoke determinism, it seems that we can generate worries about whether people are free, even in the sense of being able to do otherwise. Thus, it is not clear that any scorekeeping confusion *does* occur when we raise the parameters and explicitly ask the compatibility question about agentic experience. We may simply be exhibiting our competence in applying the notion of freedom in that context as well.

I suggest that people's incompatibilist judgments about their experience are at least *prima facie* evidence of the actual nature of the phenomenal content with which people are presented when they attend to their agentic experience. Thus, we would need some positive reason, other than the hypotheses canvassed by Hor-

worry whether Black really functions in the same way as determinism, since libertarians deny that Black *does* block alternative possibilities (even though they grant that determinism blocks such alternatives). However, even if Black does not block all alternatives, he plausibly blocks the sorts of alternatives that would be required for the ability to do otherwise. After all, many have argued that even if certain alternatives in a Frankfurt-case offer Jones a "flicker of freedom" (Cf. Fischer 1994, Ch. 7), they are not robust enough to underwrite the sort of freedom to do otherwise that might plausibly be required for moral responsibility. The central point is that while alternative possibilities (of some sort, i.e., without begging the question either way regarding the compatibility issue) are necessary for the ability to do otherwise, they are nevertheless not sufficient for such an ability. Conversely, a compatibilist might insist that determinism allows for a compatibilist ability to do otherwise, even if Black does not. That may be right. However, most compatibilists accept that determinism *does* block alternative possibilities, and thus the ability to do otherwise. Although I happen to be sympathetic to the idea that it does not, it should be noted that there is wide agreement that compatibilist accounts of such an ability are subject to fatal criticisms (Cf. Lehrer 1968; see also Clarke 2009 for why more recent compatibilist accounts of the ability to do otherwise fail).

¹⁶ In Chapter 5, I suggest another sort of contextualist move that is more plausible, and which can deal with Frankfurt-cases.

gan, to think that people systematically misinterpret the nature of such experience. Of course, if Horgan were right that judgmental introspection is unable to provide a reliable answer to the experience-compatibilism question, then we would indeed need an error theory to account for why people make the judgments that they do. In that case, Horgan's contextualist hypothesis would be preferable to the alternative hypothesis proposed here, namely that such judgments are accurate and competent (as they simply could not be). Yet, given the considerations that count against Horgan's view, his contextualist error theory looks doubtful. That is, it is unclear whether Horgan's contextualist hypothesis fares better than the alternative hypothesis that agents are competent and in normal contexts.

3.6 An Alternative Compatibilist Proposal

In this section, I give a broad outline of an approach that I will explore more fully in the next chapter. This approach takes takes people's incompatibilist reports about their experience of freedom at face value. The trick for compatibilists, I suggest, is to argue that such experience has *two* sorts of content, and thus two associated veridicality conditions. Even granting (at least for the sake of argument) that people's introspective reports about their experience are incompatibilist and reliably latch onto the relevant phenomenology, there might still a respect in which the experience is veridical, assuming determinism.

Recall that for any experience, its phenomenal content yields a veridicality condition: the content specifies how the world must be for the experience to be veridical. If the experience has two sorts of content, then it has two associated veridicality conditions. Now take agentic experience. Perhaps our experience of being free to do otherwise has two sorts of content. According to one of these, we would have to be libertarian agents for our experience to be veridical. Yet maybe this very experience also has another sort of content, which is sometimes veridical under determinism.

Consider an analogy. On what we might call a pre-Newtonian view of colors, we experience colors as primitive properties of objects, spread out over their

surfaces.¹⁷ When we see a red apple, what is presented to us in phenomenology is that a certain object, the apple, has a certain simple property, redness, spread out over its surface. This property seems irreducible: the apple's redness does not seem, at least in phenomenology, to consist in any more fundamental property—say, a microphysical or dispositional property of the apple, or some unspecified property of the apple that plays a causal role in generating our visual experience. The apple just seems primitively *red*. David Chalmers (2006) calls this *perfect content*.¹⁸ Of course, as Newton and Galileo first saw, such a view of colors is false. Physics tells us that apples are not red (or green) in anything like the way we experience them as being. For physicists like Newton and Galileo, as well as for philosophers like John Locke, the result of this discovery was counterintuitive: *there are no colors*. Thus, all our experiences of colors are non-veridical. This not only leaves us with no way of distinguishing red from green, but with no way of distinguishing—just in terms of the veridicality conditions of our phenomenology—illusory (or hallucinatory) color experiences from experiences we normally think are veridical. As I will outline in a moment, Chalmers proposes a novel way of escaping this unsatisfactory situation.

Now consider agential experience. Let us grant, at least for argument's sake, that when we experience being free to do otherwise our phenomenology presents us (to ourselves) as possessing a certain property: an indeterministic ability. Likewise, let us grant that this property does not seem, at least in phenomenology, to be any more fundamental property, such as a microphysical or dispositional property. It just seems—experientially—that we are able to do otherwise, and in a way that requires the falsity of determinism. Call this *libertarian content*. If the phenomenal content of our agential experience is libertarian, this has the result that

¹⁷ Or spread throughout a volume (e.g., wine), etc.

¹⁸ Chalmers also calls this content *Edenic*—it is the content of experiential representations of the primitive properties instantiated in “Eden” (2006: 66). In the “Garden of Eden,” Chalmers writes, “We had unmediated contact with the world. We were directly acquainted with objects in the world and with their properties. Objects were simply presented to us without causal mediation, and properties were revealed to us in all their true intrinsic glory” (48). My “pre-Newtonian world” is Eden for colors.

all our experiences of being able to do otherwise are illusory, at least assuming determinism. Thus, it leaves compatibilists with no way of distinguishing—just in terms of the veridicality conditions of our phenomenology—illusory agentive experiences from experiences we normally think are veridical. Recall our example from earlier: you wake at night and consider switching on the light. As you lie there, you experience yourself as being free to switch on the light, or to refrain from doing so by continuing to lie there motionless. What we wanted was a way of distinguishing such a case—just in terms of the veridicality conditions of the phenomenology—from one in which, unbeknownst to you, you have been paralyzed by a drug, yet nevertheless you experience yourself as being free to switch on the light or refrain from doing so. Yet if all your experiences of being free to do otherwise are illusory because determinism is true, compatibilists cannot make sense of the idea that your experience in the first case is veridical in exactly the way it is not in the second, since the phenomenology is the same. This, too, is unsatisfactory.

Before I outline my compatibilist proposal for agentive experience, let me first outline Chalmers's proposal for color experiences. When it comes to such experiences, Chalmers argues that in addition to perfect content there is also another sort of phenomenal content that makes color experiences veridical, at least in the right kinds of cases. This sort of content allows us to differentiate illusory experiences of seeing red from experiences we normally think are veridical. Chalmers calls this second sort of content *ordinary* content. Crucially, perfect content serves as a regulative ideal in picking out the ordinary content. What does this mean? For a given experience of seeing red to be perfectly veridical, we would have to live in a world in which colors actually are primitive properties of objects, spread out over their surfaces. We would have to live in a world of pre-Newtonian colors. The best we can do in our world is to have the relevant experience be caused by whatever actual properties play the role that the relevant primitive properties would play in a pre-Newtonian world. Even though no property can play this role perfectly, some property (or properties) may be able to play it well enough—i.e.,

by being, as a matter of fact, the normal cause of experiences of seeing red. This condition *constitutes* the ordinary phenomenal content of experiences of seeing red. Such experiences are veridical once they are caused by whatever properties of objects normally cause them (under normal conditions). Of course, ordinary content does not yield, by itself, an adequate account of the phenomenal content of experiences of seeing colors, since it does not capture how things seem to us phenomenologically. As a result, Chalmers suggests that color experiences should be thought of as having two sorts of content: *perfect* and *ordinary* content. Each sort of content is useful for a different purpose. Perfect content captures our color phenomenology, while also serving to pick out the ordinary content by being its regulative ideal. Ordinary content allows us to make the kinds of distinctions we want to make between illusory experiences of seeing colors and experiences we normally think are veridical.

Might not a similar story be told for experiences of being free to do otherwise? Perhaps such experiences also have two sorts of phenomenal content, and two associated veridicality conditions. First, they have libertarian content, which captures how things are presented to us phenomenologically. Second, they have another sort of content as well, which allows us to distinguish cases of illusory experiences of freedom from experiences we normally think are veridical. Call this *compatibilist* content. Analogously with the color case,¹⁹ libertarian content serves as a regulative ideal in picking out the compatibilist content. What does this mean? For an experience of being free to do otherwise to be veridical according to the standard of libertarian veridicality, which is the standard associated with libertarian content, we would need to possess exactly the libertarian freedom we experience possessing. The best we can do in a deterministic world, however, is to have our experience undergirded by whatever properties actually play the role in such a world that libertarian properties would play in a libertarian world. Even though no property can play this role perfectly, some property (or properties) may play it well enough, by being, as a matter of fact, the normal undergirding

¹⁹ Note that this analogy need not be airtight; it is only meant to be illustrative.

of our experience of being free to do otherwise. This condition *constitutes* the compatibilist phenomenal content of experiences of freedom. Such experiences are veridical once they are undergirded by whatever properties normally undergird them (under normal conditions). Of course, compatibilist content is not, by itself, an adequate account of the phenomenal content of experiences of being free to do otherwise, since it fails to capture how things seem to us phenomenologically: our experience is, we are granting *arguendo*, *as of* our possessing an indeterministic freedom. Thus, my proposal is that experiences of freedom have two sorts of phenomenal content: libertarian and compatibilist. Each sort of content is useful for a different purpose. Libertarian content captures our phenomenology, while also serving to pick out the compatibilist content by being its regulative ideal. Moreover, compatibilist content enables us to make the distinctions we want to make between illusory experiences of freedom, and experiences we normally think are veridical.

3.7 Conclusion

The proposal I have sketched needs to be worked out in greater detail, which is a task I undertake in the next chapter. Yet even in rough outline it exhibits some attractive features. First, on the assumption that there *are* agentive experiences, and specifically experiences of being free to do otherwise, even if such experiences have libertarian content there is still wiggle-room for compatibilists to argue that these very experiences are veridical, assuming determinism. This enables compatibilists to make the sorts of distinctions that they need to make between illusory experiences and veridical experiences, while assuming determinism. Second, my view opens the way for a compatibilist theory of freedom to capture the ability that most people apparently experience having, even if such experience has libertarian content. Finally, my proposal blocks the libertarian move of trying to justify belief in libertarianism on the basis of people's experiencing being indeterministically free. On my view, there is no more reason to endorse libertarianism

on the basis of our experience than there is to endorse compatibilism. Indeed, my view has a further attractive feature, since it concedes that libertarianism is at least *descriptively* right about the content of phenomenology. Thus, my proposal goes further than Horgan's in addressing incompatibilist concerns while nonetheless remaining *prescriptively* compatibilist. By all means, compatibilists may now grant, our agentive experience has content that is non-veridical under the assumption of determinism, just as libertarians claim. Yet that does not pose any problem for compatibilism.

In closing this chapter, I want to note the following. The view that I have sketched here (and which I develop further in Chapter 4) differs from Horgan's in a crucial respect. My view can grant that people's experiences are descriptively incompatibilist (as the evidence adduced in Chapter 2 suggests it is), and thus are in error in the sense of being (in part) non-veridical if determinism is true. By contrast, Horgan's view says that people's experiences are *not* descriptively compatibilist, and so they are not even partly non-veridical under the assumption of determinism. The error, Horgan thinks, occurs not in the experience's being non-veridical, but rather in the judgments that people make about the experience. So, in positing a gap between the descriptive and prescriptive questions, I *am* committed to the possibility of something's being mistaken or in error if prescriptive theory says that how we *should* think about modal freedom differs from how we *do* think about it. Yet that does not mean that I am committed to an error theory in the sense that Horgan is.

Chapter 4

Against an Argument for Libertarianism

Libertarians often claim that our experience of freedom is indeterministic. This claim then functions as a step in an argument in favor of libertarianism, the view that freedom requires indeterminism and we are free. Since, all else being equal, we should take our experience at face value, libertarians argue, we should believe libertarianism. I argue that compatibilists, who think that freedom is consistent with determinism, can do better than their usual responses to this argument. I begin by conceding that our experience of freedom is in a sense libertarian, but I argue that it is also in another sense compatibilist. Thus, even if libertarian descriptions of the experience are correct, there is still a sense in which that very experience might be veridical, assuming determinism.

4.1 Introduction

This chapter responds to a certain argument in favor of libertarianism, the view that freedom requires indeterminism and we are free:

Argument L

- (1) The content of our experience of freedom is (presumptively) veridical.¹
- (2) Our experience of freedom has libertarian content.
- (3) If our experience of freedom has libertarian content, then our experience is veridical only if libertarianism is true.
- (4) So, libertarianism is (presumptively) true.²

The libertarian content is this. For a possible action, A, that one is considering performing, one's experience is as of being able to decide to A, and as of being able to decide, in an unconditional sense,³ to refrain from A-ing.

For compatibilists, who maintain that freedom is consistent with determinism, the obvious responses to Argument L are to reject (1) or to reject (2). This chapter makes a case for rejecting (3), and also for the claim that (1) is neither true nor false since it presupposes that there is a unique content to experiences of freedom.

4.2 Compatibilist Strategies

Some compatibilists think that premise (1) of Argument L is false.⁴ These compatibilists grant premise (2), and thus admit that the experience of freedom has

¹ More will be said later about veridicality, but for now let us loosely define the term 'veridicality' as: accurate in a relevant sense, user some specification of how the world might be. Different ways of spelling out what exactly this means will be sketched later, but this broader, univocal sense of 'veridicality' will remain constant throughout the chapter.

² This argument appears, for instance, in Reid (1788), C. A. Campbell (1951: 463), and O'Connor (1995: 196–197).

³ According to the compatibilist "conditional analysis" of the freedom to do otherwise, an agent, S, is free to do otherwise iff the following is true: If S wanted (or tried, etc.) to do otherwise, then she would do so. That is, something about S's beliefs, desires, and so on, would have to be different in order for S to do otherwise. Libertarians think that S is free to do otherwise given her actual beliefs, desires, etc. This is what is meant by saying that S is free to do otherwise "in an unconditional sense"; it is possible that S will do otherwise given her actual mental states. Each of the compatibilist's and incompatibilist's claims usually require the laws of nature to be the actual laws.

⁴ For example, Lehrer concedes that when it comes to experiences of freedom, the libertarian "accurately describes what I find by introspecting, and I cannot believe that others do not find the

libertarian content, but they insist that it is non-veridical. However, compatibilists who adopt this strategy concede what arguably they should not concede: that our experience of freedom is non-veridical if determinism is true. Adopting this strategy also requires showing that we lack libertarian freedom, rather than merely showing that there is a variety of freedom that is compatible with determinism. It is not clear that compatibilists need to take on this extra argumentative burden.

Most compatibilists prefer to reject (2).⁵ These compatibilists insist that our experience of freedom is veridical, but they claim that it has exhaustively compatibilist content. The experience is simply of having a conditional freedom: as long as we are free from constraint, coercion, and so forth, our experience is as of being free to decide to A, or to refrain from A-ing, *if* we want (or try) to do so.⁶ By rejecting (2), though, compatibilists enter into intractable disputes with libertarians⁷ about the nature of the presentational content of our experience of freedom—about whether the content is of having a compatibilist (conditional) freedom or instead a libertarian (unconditional) freedom. It is exceedingly difficult to know how to adjudicate such disputes, given that they turn on competing introspective claims. Is one side mistaken about what it introspects?

With this in mind, another compatibilist strategy is to develop an error theory for libertarian judgments about experiences of freedom. According to this strategy, even if people judge their experience as incompatibilist, actually it is compatibilist: people misinterpret their experience.⁸ Dialectically, this is more

same” (1960: 150). So Lehrer concedes (2). Yet he thinks that our experiences of freedom are not veridical, and thus he disables Argument L at step (1). Similarly, Hume thinks that “There is a false . . . experience . . . of the liberty of indifference” (1960/1739: Bk. II, Part III, §II), by which he means that our experience of freedom is libertarian in nature, yet is non-veridical.

⁵ See, for instance, John Stuart Mill (1865: 285), Adolf Grünbaum (1952: 672), and Terry Horgan (e.g., 2012; 2011), the latter of whose views I discussed in Chapter 3, and I discuss again in a moment.

⁶ See footnote 3 above regarding the compatibilist “conditional analysis” of freedom.

⁷ And, more generally, with incompatibilists, who think that freedom is incompatible with determinism.

⁸ Horgan (e.g., 2012; 2011) makes this compatibilist move. See Chapter 3 for details.

helpful than just banging one's fist on the table and insisting that the experience is compatibilist. Even so, it commits one to arguing that the experience is *exhaustively* compatibilist. It would be useful if compatibilists could avoid shouldering this burden.

This paper presents a case for rejecting (3). The trick for compatibilists is to argue that our experience of freedom has two sorts of phenomenal content, and thus two associated veridicality conditions. The experience might be veridical if the satisfaction conditions of just one of these types of content are met. If one type of phenomenal content is libertarian, yet an experience of freedom has another type of phenomenal content that is compatibilist, the experience might be veridical even if only the latter content is satisfied. Libertarianism need not be true. Thus, Argument L would be blocked at premise (3). However, since libertarianism plausibly implies the view that libertarian phenomenal content is veridical, (1) and (2) jointly entail (3). Thus, (3) cannot be denied without denying (1) or (2). In the end, this chapter makes the case that (1) is neither true nor false since it presupposes that there is a unique phenomenal content to experiences of freedom.

4.3 An Analogy with Color

An experience that has libertarian content might be veridical if determinism is true. This is because such an experience might have more than one distinct type of phenomenal content. The experience might be veridical if the satisfaction conditions of just one of these types of content are met. Here, “phenomenal content” is content that is tied in a particular way to an experience's phenomenology: it is constitutively determined by the experience's phenomenal character.⁹ We can

⁹ This is not to beg any questions about content determination, an important issue in the philosophy of mind. In short, we can be neutral about whether content or phenomenology has explanatory priority. According to some philosophers, we should explain phenomenology in terms of content: in virtue of redness being represented in experience, for instance, one has phenomenology as of redness being instantiated. So, intentional properties constitutively determine phenomenal

see how this might work by first looking at how David Chalmers (2006) makes a similar move in connection with the phenomenal content of color experience.

Chalmers thinks that the view about phenomenal content that is most adequate to our phenomenology in visual color experiences is *primitivism*. On this view, we experience colors as simple intrinsic properties of objects, spread out over their surfaces. As Chalmers puts it,

When I have a phenomenally red experience of an object, the object seems to be simply, primitively, *red*. The apparent redness does not seem to be a microphysical property, or a mental property, or a disposition, or an unspecified property that plays an appropriate causal role. Rather, it seems to be a simple qualitative property, with a distinctive sensuous nature. (2006: 66)

If this is right, then experiences of color have contents that attribute primitive properties. This is primitivist content.¹⁰ Chalmers thinks it natural to consider such content to be phenomenal content, given that the properties presented in the experience seem to be fully determined by the phenomenology.

However, Chalmers concedes that, “For all its virtues with respect to phenomenological adequacy, the . . . primitivist view has a familiar problem. There is good reason to believe that the relevant primitive properties are not instantiated in our world” (2006: 66). According to primitivism, then, none of our experiences of color is veridical. As a consequence, primitivism fails to provide

properties. (Proponents of such a view must then explain how content *suffices* for phenomenology; see e.g., Tye (1995: §5.2), Lycan (1996), Rosenthal (1990).) By contrast, others maintain that we should explain content in terms of phenomenology: in virtue of the property of redness being presented in phenomenology, one’s experience has content that derives from this property. So, phenomenal properties constitutively determine intentional properties. (Defenders of this view are less numerous; they include Horgan and Tienson (2002), and Kriegel (2002).) The reason we can be neutral on the question of explanatory priority is that even if one’s experience has its content in virtue of what is presented in phenomenology, it might still be true that one has such phenomenology in virtue of first having content (Cf. Chalmers (2006: 51)). Presumably, one’s having that content would then be explained by an intentionality-first theory, plus a story about how mental states get content. So, we can leave it open whether content determines phenomenology, or vice versa, while still making use of the notion of phenomenal content.

¹⁰ See e.g., John Campbell (1997) for a canonical statement of this sort of view.

us with any way of distinguishing—just in terms of the veridicality conditions of our phenomenology—illusory or hallucinatory color experiences from experiences that are accurate.

In addition to primitivist content, Chalmers argues that there is another type of phenomenal content that makes color experiences accurate, at least in the right kinds of cases. This second content enables us to differentiate illusory or hallucinatory experiences from experiences we know to be accurate. This is ordinary content, which has its own associated veridicality condition: it is satisfied iff the relevant object has whatever property (or set of properties) normally causes phenomenally red experiences.¹¹

Ordinary content is not adequate to our phenomenology, since it does not reflect the phenomenal character of our color experience.¹² As a result, while this second content reflects our judgments about veridicality, it fails the important test of phenomenal adequacy, which primitivism passes.

Chalmers's idea is to combine these two views in a way that captures both the truth-conditional virtues of ordinary content, and the phenomenological virtues of

¹¹ The central idea behind this sort of view is that the attribution of color concepts is justified, and our experience of colors is veridical, despite the evident fact that primitive color properties are not instantiated in our world. See, for instance, Johnston (1992) for an influential argument to the conclusion that part of what a philosophical theory of color should do is to make sense of the situations and practices in which we apply color concepts. Johnston grants that in an important sense, our world is not colored, since the driving belief behind primitivism, which is that the intrinsic nature of a color is fully revealed by visual experience as of a given colored object, is most likely false, or at least incompatible with other core beliefs we have about color. Johnston calls this belief "Revelation" (1992: 223). Even so, he argues that in another important sense, our world *is* colored, since there are (usually) properties instantiated that make true "enough" of our beliefs about color. Johnston's view is that colors are best conceived as dispositional properties. Chalmers denies this claim (2006: 56–58), while nevertheless adopting something close to Johnston's strategy. In place of Johnston's dispositionalism, Chalmers endorses what he calls a "quasi-Fregean" view (2002: 135) about how color terms refer, and about how the content of color experiences is veridical (2004, 2006) despite its not being veridical in the way that it *would* be were primitive color properties actually instantiated in our world.

¹² This is because, as Chalmers puts it (in the passage quoted above), a phenomenally red experience is as of a "simple qualitative property, with a distinctive sensuous nature," and not as of "a microphysical property, or a mental property, or a disposition, or an unspecified property that plays an appropriate causal role" (2006: 66).

primitivist content.¹³ His proposal is this. For color experiences to be *perfectly* veridical, objects would have to instantiate primitive color properties. Even if an experience is not veridical in this way, however, it might nevertheless be *imperfectly* veridical: it might be veridical according to our ordinary standards of accuracy.¹⁴ These are the standards according to which we differentiate (in everyday life) veridical from non-veridical experiences of colors—as when we see, rather than merely hallucinate, a red apple. According to Chalmers, there is no conflict here, once we keep in mind that the two notions of veridicality are associated with distinct *conditions* of veridicality. Imperfect veridicality is associated with an ordinary veridicality condition, while perfect veridicality is associated with a primitivist condition. The result is that color experiences have more than one type of phenomenal content, depending on the associated notion of veridicality, and an experience is veridical as long as one of these is satisfied.

The most fundamental type of content, Chalmers argues, is primitivist content. This is because primitivist content *determines* ordinary content via a “matching” relation, which works as follows. For a color experience to be perfectly veridical, we would have to live in “Eden,” which is a world in which primitive color properties are instantiated by objects. The best that we can do in our world is to have certain properties “match” the primitive properties attributed by primitivist content, by playing the role that these properties *would* play in Eden. While no property can play this role perfectly, some property (or properties) may be able to play it well enough, by being the normal cause of phenomenally red experiences. So, ordinary phenomenal content is grounded in primitivist phenomenal content. In Chalmers’s terms, primitivist content serves as a “regulative ideal” in determining the ordinary content. That is, the primitivist content sets an ideal standard for the veridicality of phenomenal content, and the ordinary phenomenal content is

¹³ This is why Johnston’s view (1992; see footnote 11 above, is an important precursor to Chalmers’s strategy.

¹⁴ As Chalmers notes (2006), it does not follow that merely because an experience is imperfectly veridical, it is not “really” veridical. If anything, it seems more plausible that imperfect veridicality is what our ordinary term ‘veridicality’ picks out.

a condition that relates us to whatever properties in fact come closest to meeting this standard.

A similar story can be told for experiences of freedom. Even if the presentational phenomenal content of such an experience is libertarian, and therefore is non-veridical if determinism is true, there is still a second type of phenomenal content that might be veridical. This second content is compatibilist.¹⁵

Let us frame this strategy in another way. It is an empirical discovery that primitive color properties are not instantiated in our world. As a result of this discovery, we are left with two options. First, we can conclude that *the* content of color experiences is non-veridical. The result of this would be that none of our experiences of color is veridical. Second, we can rescue the intuition that there is still a sense in which our color experiences *are* veridical, by introducing ordinary content. Similarly, the claim of the present chapter is that the correct response to Argument L is to say that it is an empirical question whether (1) and (2) are jointly true. At best, Argument L establishes a defeasible presumption in favor of libertarianism. If it turns out that libertarianism is empirically defeated (e.g., if indeterminism turns out to be false), then even if we grant, *arguendo*, that experiences of freedom have libertarian content, we can still rescue the intuition that there is a sense in which experiences of freedom are veridical, by appeal to compatibilist content.

4.4 Libertarian Content

Assume that there is something it is like to be an agent, and that agentive experiences have rich contents. In perceptual experience, rich contents attribute not only low-level properties like redness or squareness, but also high-level properties like being an apple. Likewise, rich contents for agentive experience attribute high-level properties like being free to do otherwise, in addition to low-level properties like being an action.

¹⁵ Pereboom (2011: 29–40) uses Chalmers’s theory in developing a account of phenomenal concepts. The view presented here is indebted to Pereboom’s use of Chalmers’s theory.

One might doubt whether there are experiences of freedom. Moreover, even if one grants that there are such experiences, one might doubt whether they can have content that is rich enough to be non-veridical if determinism is true.¹⁶ However, let us grant for the sake of argument that this idea makes sense. Assume that premise (2) of Argument L is true: there are experiences of freedom that have libertarian phenomenal content.

This phenomenal content has at least two parts. For some action, A, the relevant phenomenal content is that (i) one is able to decide to A, and (ii) whether one decides to A is not determined. Here, (i) is agentive. Yet (i) is not enough for the content to be libertarian. We need (ii) for that.

How could anyone's experience have as content that her deciding to A is undetermined by her prior states together with the laws of nature?¹⁷ After all, our experience presumably does not concern such laws. Furthermore, even if someone's experience is *not* that her deciding to A is determined by her prior states (together with the laws), it would clearly be a mistake to conclude from this that her experience is thereby that her deciding to A is *not* determined by those states. For present purposes, however, let us assume that content (ii) is of one's feeling a certain unconditional openness to the future. The future feels open in a way that would require indeterminism for the feeling to be accurate. It feels "as if" one is free to decide to A or, in an unconditional sense, to refrain from A-ing.

We are assuming for the sake of argument that experiences of freedom have libertarian content of the sort just described. This is what is premise (2) of Argument L says. By analogy with primitivism about color experiences, it may seem that an experience with libertarian phenomenal content is veridical iff libertarianism is true. Call a world in which libertarianism is true an "Agentive Eden." Perhaps the actual world is an Agentive Eden, although very likely it is not. Leave that question aside. What concerns us here is only whether experiences of freedom

¹⁶ After all, the richer the content is, the more demanding the veridicality condition will be.

¹⁷ The standard characterization of determinism is that whether one decides to A is entailed by a description of one's prior states together with a statement of the laws.

with libertarian content might be consistent with determinism, in the sense of being veridical even if determinism is true.¹⁸ This question can be answered without coming to any verdict about whether libertarianism is true or instead determinism is true.¹⁹

On the face of it, a libertarian experience of freedom is veridical only if we live in an Agentive Eden, and so it would seem to be obviously non-veridical in a deterministic world. However, the proposal to be outlined in the next section says that this is not obvious. An experience of freedom with libertarian content might be veridical, even assuming determinism (and thus despite the fact that the libertarian content is not veridical).

4.5 Compatibilist Content

By analogy with the color case, the proposal of this paper is that there is a second sort of phenomenal content to experiences of freedom that is compatibilist. This content is a condition that a property must satisfy in order to be the property that is attributed by the experience.²⁰ The property attributed by the experience is, of course, the freedom to do otherwise. What condition might work as the second content for such an experience?

¹⁸ Note that when we ask whether someone's experience as of being able to do otherwise is compatible with determinism, we are asking whether her experience is veridical, assuming determinism. We are not asking whether the experience itself is compatible, even if it has libertarian content. After all, presumably any experience whatever, including a libertarian one, is potentially compatible with determinism, since we could be determined to have that very experience (Cf. Mele 1995: 133–37, 246–49).

¹⁹ Obviously, one cannot determine whether a given content is veridical without specifying an evaluation context, which is either deterministic or not. Even so, one does not have to be committed regarding either (a) whether any experience actually has that content, or (b) whether the specified evaluation context is actually true. One is merely assuming both the content and the evaluation context in order to assess whether that content could be veridical in that context, in the same way that one can assess whether two claims are compatible without being committed to the truth of either claim.

²⁰ According to Chalmers's "quasi-Fregean" view, the phenomenal content of an experience is the mode of presentation of a property presented in that experience, not the property itself. This is what Chalmers calls an "epistemic intension" (2002: 135). For experiences, such an intension is "a condition that a property must satisfy in order to be the property attributed by the experience" (2006: 59).

For color, the second content is *whatever property (or set of properties) ordinarily causes phenomenally red experiences*. In the agentive case, however, we cannot say that the second content is *whatever property (or set of properties) ordinarily causes experiences of being free to do otherwise*, since presumably no one thinks that the meaning of ‘being free to do otherwise’ is *being such that it causes an experience of being free to do otherwise*. More plausibly, the second content is the following condition: that there is instantiated whatever relevant property (or set of properties) is ordinarily instantiated when one experiences being free to do otherwise. This content, which is compatibilist, is veridical iff this condition is met.

There is good reason to think that this condition really does pick out a genuine second content for experiences of freedom. Consider that it is at least somewhat plausible to think that ‘freedom’ is a natural-kind term that refers to whatever relevant processes are at work in choices or decisions that we ordinarily call ‘free,’ or that feel free to us. Consequently, we might be free even if determinism is true.²¹ It follows from this view that so long as the relevant processes actually constitute a natural kind, we are free.²²

Another way to state this sort of compatibilist position is as follows. We often begin theorizing about freedom by pointing to paradigm cases—for example, cases in which an agent acts while free from coercion, constraint, and so forth. The paradigm-case view says that we should stop there: simply point to actual choices or decisions that we call ‘free,’ or that feel free to us, and those are paradigm cases of freedom.²³

²¹ Heller (1996) has argued that ‘freedom’ (or ‘free will’) is plausibly a kind term comparable to natural-kind terms. Heller thinks that in just the way that cats might turn out to be non-biological robots, so our free actions may turn out to be determined.

²² Cf. Heller (1996: 336). If the choices we call ‘free’ do not constitute a kind, then there are no free choices.

²³ The paradigm-case view is suggested by Flew (1959), among others. For criticism of the view, see Danto (1959), van Inwagen (1983: ch. 4), and Double (1996: ch. 2; 1997). As Heller points out (1996: n.7, 336), the paradigm-case view is actually stronger than the kind view about freedom, since it *guarantees* that there are free choices, whereas the kind view does not (see footnote 22).

The natural-kind view is not unproblematic. One problem is that ‘freedom’ seems to behave differently from standard natural-kind terms like ‘water’ (cf. Balaguer 2010: n.5, 22).²⁴ ‘Water’ refers to the aqueous stuff that flows in our rivers and falls from our sky as rain, whatever that stuff happens to be. If we were to discover that our aqueous stuff is actually XYZ rather than H₂O, then ‘water’ would refer to XYZ (and would have done so all along). One thing that presumably we would not say if we made such a discovery is that water does not exist. The problem for the natural-kind view about freedom is that ‘freedom’ appears to work differently from ‘water’ in this regard. If we were to discover that the processes at work in choices that feel free to us include our being remotely controlled by Martians, it seems that we would not conclude from this that ‘freedom’ refers to our being remotely controlled by Martians. If we made such a discovery, it would be a discovery that we lacked freedom. So the natural-kind view about freedom seems mistaken.

This problem might be remedied. Imagine that what we experience as water is, as it turns out, a hologram accompanied by remotely-induced sensations. There is no aqueous stuff at all. Nevertheless, it might be true that ‘water’ refers to a natural kind just so long as no hoax of this sort is going on. The same goes for freedom. As long as nothing like remote control by Martians is going on, ‘freedom’ refers to a natural kind.

Even granting that Martian-control scenarios do not threaten a compatibilist natural-kind view about freedom, incompatibilist arguments put pressure on such a view by suggesting that libertarian conditions on freedom should be included among the constraints that must be satisfied for there to be a proper referent for ‘freedom.’ In order to provide a non-question begging defence of the claim that ‘freedom’ picks out a compatibilist natural kind, the appeal of such incompatibilist arguments must be explained away.²⁵

²⁴ Mark Balaguer talks about ‘free will.’ ‘Freedom’ picks out the same thing as ‘free will.’

²⁵ I owe this point to Gunnar Björnsson, in correspondence.

Whatever intuitive plausibility the natural-kind view about freedom has, the analogous view for *experiences* of freedom is more plausible. This is because the analogous view is part of an overall position that already accommodates an incompatibilist perspective. By granting libertarian phenomenal content, the claim that there is a second, compatibilist content for experiences of freedom does not come under the same sort of pressure from incompatibilist arguments as the natural-kind view about freedom does. The analogous view says that one type of phenomenal content is libertarian, and is non-veridical under the assumption of determinism. Still, such experiences have a second type of phenomenal content that is compatibilist. This content is veridical under normal conditions, even assuming determinism.

What are normal conditions? First, they are conditions in which the relevant experience is not obviously illusory. Consider an example. Someone has been hypnotized. A performing hypnotist has primed a subject to pour a glass of water over her own head as soon as she hears the hypnotist cough. The subject is unaware of this. The hypnotist tells his audience what will happen. Then he gives his subject a glass of water. He tells her that if she drinks it in less than five seconds, she will win a prize. He also reminds her that she should not feel obliged to drink it. The subject agrees. She takes the glass, the hypnotist coughs, and the subject pours the water over her own head.

When the subject accepts the glass of water, presumably she experiences being free to do various things: to drink it, or to refrain from drinking it (e.g., by refusing it, or by pouring it over her own head, or . . . etc.). Yet the hypnotist is skilled: whenever he succeeds in hypnotizing a subject, the subject never fails to pour the water over her own head. In this case, too, the subject is unable to do otherwise. If so, then her experience of being free to do otherwise is non-veridical, and obviously so.²⁶

²⁶ This experience is illusory on both contents—libertarian and compatibilist. The point is that unless we introduce compatibilist content, then if indeterminism is false we have no way of distinguishing (just in terms of the veridicality conditions of the experience) non-veridical experiences

Call this a *locally* abnormal situation. Experiences of freedom that are not illusory in anything like this way are candidates for those that occur under normal conditions.

Additionally, we might want to rule out *globally* abnormal situations, such as everyone's being remotely controlled by Martians. Thus, we might add a requirement that there are certain constraints that must be satisfied—for instance, our not being controlled by Martians—in order for a situation in which we experience being free to do otherwise to count as normal.

In normal situations, then, one's experience of freedom is veridical iff a certain condition is met. This condition is that one instantiate whatever relevant property (or set of properties) is ordinarily instantiated when one experiences being free to do otherwise. This is compatibilist content.

4.6 The Two-Stage View

How are the two types of phenomenal content related? Among them, the most fundamental type of content is libertarian. This is because we are assuming that libertarian content is the content that most closely reflects our presentational phenomenology in experiences of freedom: it accurately reflects (we are assuming) what is presented in our phenomenology.²⁷

Analogously with the color case, presentational content determines the second content via a matching relation. For an experience of being free to do otherwise to be perfectly veridical, we would have to live in an Agentive Eden. The best that we can do if determinism is true is to have certain properties match the libertarian properties that are attributed by the presentational content, by playing the role that these properties *would* play in an Agentive Eden. No property can play this role

like this from experiences we normally judge to be veridical.

²⁷ If premise (2) of Argument L were false, and experiences of freedom actually had exhaustively compatibilist content, then our presentational content would be compatibilist.

perfectly. Yet some property (or set of properties) may be able to play it well enough, by being the property (or set of properties) that is ordinarily instantiated when one experiences being free to do otherwise. Thus, compatibilist content is grounded in libertarian content. Or, to use Chalmers's phrase, presentational content acts as a "regulative ideal" in determining the second content. That is, the libertarian phenomenal content sets an ideal standard for veridicality, and the second, compatibilist content is a condition that relates us to whatever properties in fact come closest to meeting the ideal standard. Once the second content is satisfied, the experience is imperfectly veridical, even if determinism is true.

Despite granting premise (2) of Argument L, which says that the relevant presentational content is libertarian, our experience of freedom might be veridical even if it turns out that determinism is true.²⁸ Thus, Argument L is disabled at its third premise. The experience can be veridical, it can have libertarian con-

²⁸ It might be objected that if there are two phenomenal contents to our experience of freedom, then the claim 'our experience of freedom is veridical' is ill-defined, since veridicality is judged relative to a content and context, yet this claim only gives the context, not the content. However, the present chapter adopts Chalmers's pluralism about representational contents, which amounts to the claim that there are multiple content relations (i.e., relations that associate experiences with contents). If one relation associates a given experience (such an experience of freedom) with one sort of content, nevertheless another relation might very well associate *that* experience with another sort of content. In this way, an experience of freedom might be associated with more than one sort of content via different content relations. Of course, as Chalmers notes (e.g., 2006: 51–52), not all of these contents can be *phenomenal* contents, since if a content varies independently of the phenomenal character of the relevant experience, then that content is not phenomenal content. So, even if it is definitional of an experience that it has phenomenal character, still the contents of an experience do not have to be exhausted by the content associated with its phenomenal character. However, here we are focusing just on an experience's *phenomenal* content relations, and thus the claim is that a given experience—for instance, an experience of freedom—might well have more than one phenomenal content relation and, as a result, more than one sort of phenomenal content. As Chalmers puts it, "For ease of usage, I will speak of *the* phenomenal content of an experience, but we should leave open the possibility that there is more than one phenomenal content relation, so that a given experience can be associated with phenomenal contents of more than one sort" (2006: 52). Thus, the claim of the present chapter is that (for ease of usage) *an* experience of freedom that is associated with libertarian phenomenal content via one content relation, might well be associated with another, compatibilist phenomenal content via another content relation, such that even if (for ease of usage) *it* is non-veridical when assessed according to its libertarian content (e.g., if determinism is true), still *it* might be veridical when assessed according to its compatibilist content.

tent, yet libertarianism can be false. Since libertarianism plausibly implies the view that libertarian phenomenal content is veridical, however, premises (1) and (2) of Argument L jointly entail (3). Thus, we cannot deny (3) without denying (1) or (2). This chapter has argued that (1) is neither true nor false since it presupposes that there is a unique phenomenal content to experiences of freedom, namely libertarian content. Yet experiences of freedom plausibly also have the sort of compatibilist content just defended. If we discover that libertarianism is false because determinism is true, then even granting (arguendo) that experiences of freedom have libertarian content, there is still a sense in which experiences with libertarian content are veridical (despite their veridicality not being in virtue of the libertarian content).

4.7 Conclusion

That is the case for rejecting premise (3) of Argument L, and thus Argument L itself.

The unsoundness of Argument L undermines a central motivation for libertarianism, since it removes any presumption in favor of libertarianism based solely on our experience of freedom. Showing (3) to be false also secures new ground for compatibilism. Indeed, this strategy has theoretical advantages over rival compatibilist responses to Argument L.

First, responses that grant libertarian phenomenal content by granting premise (2) of Argument L, yet which deny the presumed veridicality of the experience by denying (1), concede what they should not concede: that our experience of freedom is entirely non-veridical if determinism is true. The compatibilist view defended here makes the experience veridical, despite granting that it has libertarian content. The verdict about premise (1) is that it is neither true nor false, since it presupposes a unique phenomenal content to experiences of freedom—namely, libertarian content. The central claim of this chapter is that this presupposition is false.

Second, responses that deny premise (2), and which thereby insist that the

relevant experience has exhaustively compatibilist presentational content, cannot avoid entering into intractable disputes with libertarians about the nature of the presentational content of experiences of freedom. By accepting, at least for argument's sake, that the presentational content is libertarian, the strategy proposed here avoids such disputes. *Let* such content be libertarian. That is no threat to compatibilism.

Chapter 5

Indeterminism, Experience, and Compatibilism

Recent evidence shows that (1) people tend to believe that they possess indeterminist freedom, and (2) people also experience possessing such freedom. Some also argue that (3) belief in indeterminist freedom has its source in people's experience. Shaun Nichols denies (3), despite endorsing (1) and (2). Nichols provides an alternative account of the source of indeterminist beliefs. I argue that Nichols' account has significant shortcomings, and that belief in indeterminist freedom has (in part) its source in indeterminist experience. I explain how this works by appeal to the phenomenon of prospection, which is the mental simulation of future possibilities. Crucially, prospection can be experienced. Further, because of the way in which prospection models choice, it is easy both to experience and to believe that one's choice is indeterministic. Even so, belief in indeterminist freedom is not justified.

5.1 Introduction

In this chapter, I address the question where the libertarian content of agentive experiences of modal freedom comes from. As we saw in chapters 3 and 4, even if

people actually have experiences of being free to do otherwise, it remains unclear whether such experiences could possibly have content that is rich enough to be non-veridical if determinism is true. In chapters 1–4, we also saw that libertarians tend to think that people’s experience of deliberating and choosing actually has libertarian content. Moreover, libertarians think that this sort of experience of freedom leads people to believe that human choice is indeterministic. (The results of the studies I reported in Chapter 2 indicate that libertarians may be right about this.)

In previous chapters, I assumed for the sake of argument that the relevant libertarian content was of one’s feeling a certain unconditional openness to the future: the content is of the future being open in a way that would require indeterminism for the experience to be accurate. Where might this sort of experiential content come from? I also assumed that people’s beliefs about freedom plausibly have their source in experience, and thus libertarian beliefs can be explained by appeal to libertarian experience.

Shaun Nichols (2012) argues that belief in libertarian freedom cannot have its source in experience. Nichols argues instead that people believe that they possess indeterminist or libertarian freedom since they think that the psychological factors that are introspectively accessible do not determine their choice. Since people believe that they have introspective access to all the relevant factors that influence their choice, they infer that their choice is not determined. But this inference is not warranted, since people do not actually have access to all the factors that influence their choice.

As I will argue in the present chapter, a shortcoming of Nichols’ view is its backward-looking focus: it relies on the idea that people’s indeterminist beliefs derive from introspection on the causes of decisions. By contrast, I will argue that people’s indeterminist beliefs derive from their experience of navigating into the future, rather than from introspection on the causes of decisions.

A number of recent authors have embraced a forward-looking model of agentive experience. They think that people's agentive experience enables them—minimally—to distinguish self-generated actions from both involuntary bodily movements and the externally caused movements of objects. Tim Bayne (2011, 2008) develops a view of this sort. Yet Bayne's view is unable to explain people's experience of indeterminist freedom, and so it cannot explain how people's belief in indeterminist agency derives from their experience as deliberating agents. This is because Bayne's model focuses on perceptual experience, and it is difficult to see how such experience could have as content that one is free to do otherwise. That would require a comparison of two or more distinct representations—the alternative possibilities themselves—in the mind, and that is not a perceptual operation.

I explain people's experience of indeterminist freedom, and thus the source of their belief that they possess such freedom, by appeal to *prospction*, which is the mental simulation of future possibilities for the purpose of guiding action. Crucially, *prospction* can be experienced, and because of the way in which *prospction* models choice, it is easy for deliberating agents to experience (and to think of) their alternatives as ones they can “get to” in the sense that libertarians maintain might be provided for by indeterminism. Even so, belief in indeterminist freedom is not thereby justified.

In this way, my view provides a deflationary explanation of a major motivation for libertarianism—namely, the practical experience of deliberation and choice—while nevertheless granting that it is natural for agents to believe that their choice is indeterministic. At bottom, people's libertarian beliefs derive from their experience of navigating into the future, rather than from introspection on the causes of their choices. Thus, people's indeterminist beliefs have their source in the experience of deliberating and choosing, exactly as libertarians maintain, and against Nichols' claim.

5.2 Do Indeterminist Beliefs Have Their Source in Experience?

Many people apparently believe that human choice is (at a minimum) undetermined. They believe that when it comes to human choice and decision-making, the world is, as John Martin Fischer puts it (borrowing a phrase from Jorge Luis Borges), a “Garden of Forking Paths.” In Fischer’s words:

We naturally think of the future as open. We think of the future as containing various paths that branch off one past; although we know we will travel along just one of these paths, we take it that some of the other paths are (at least sometimes) genuinely accessible to us. In deliberating and deciding on a course of action, we intuitively think of ourselves (at least sometimes) as determining which path to take, among various paths we *could* take. (1994: 190)

If determinism is true, however, there is only one physically possible path into the future. In other words, assuming determinism, there is just one physically possible way for any agent to extend the present into the future, given the laws and the prior states of the world (Cf. Fischer 1994: 88). Thus, the way in which deliberating agents believe that they can choose among alternatives seems incompatible with determinism. If agents believe they are free in this way, they believe that they possess indeterminist freedom.

Libertarianism is the philosophical formalization of this view: freedom is incompatible with determinism, and we are free. A central element in libertarianism is the belief that indeterminism is usefully implicated in decision-making: indeterminism is necessary for free choice. Since libertarians believe that we *are* free, they consequently believe that indeterminism is true.

Libertarianism is not widely held as a philosophical view. Yet it appears to be strongly implicated in ordinary thinking about human agency, given that partic-

ipants in a number of experimental studies have tended to regard human choice as indeterministic. In one experiment (Nichols and Knobe 2007), participants were given descriptions of a deterministic universe (A) and another universe (B) in which everything is determined *except* human choices. Participants were asked “Which of these universes is most like ours?” More than 90% of respondents said that the indeterministic universe, i.e., universe (B), is most like our own (2007: 669). Further evidence in support of the view that people think of human choice as indeterministic comes from experimental studies conducted by Nichols (2004). In one experiment, children were asked whether an agent was able to do otherwise than she did, even if everything stayed exactly the same right up until the moment she made her choice. A sample question went as follows:

Joan is in an ice cream store and wants some ice cream. She chooses vanilla . . . If everything in the world was the same right up until she chose vanilla, did Joan have to choose vanilla? (2004: 486–7)

The children were old enough to understand counterfactual conditionals and were given comprehension checks to ensure they understood what was being asked. They were also given a contrast case involving a pot of boiling water, and were asked whether, with everything staying the same right up until the water boiled, it was possible that the water would not have boiled. The results showed a significant difference between the agentive case and the non-agentive case involving the boiling water. Participants tended to agree in the agentive case that the agent could have done otherwise, yet they tended to deny that the water could not have boiled. Nichols interprets these results as indicating that people regard human choice as indeterministic.

Where do indeterminist beliefs come from? Libertarians often cite experience as the source. They think that the way in which people experience their choices leads them to believe that choice is indeterministic. The results of the studies reported in Chapter 2 indicate that libertarians may be right about this. Recall that in those experiments, participants were asked to decide between two options (for

instance, two charities) and were asked whether, as they faced their decision, they experienced being able to choose either option. Participants were free to interpret such an ability however they wished. Most participants reported experiencing being able to choose either way. Determinism was then explained to the participants. Following comprehension checks, participants were asked whether the experience they had earlier reported having was consistent with determinism. Most participants tended to judge their experience as incompatibilist, whether the decision was present-focused or retrospective, imagined or actual, or morally salient or morally neutral. The only case in which participants did not give an incompatibilist response was when they were asked whether their experience of ignorance of the future was consistent with determinism. This suggests that libertarians are right in claiming that people's belief in indeterminist freedom has its source in experience.

Yet according to Nichols (2012), belief in indeterminist freedom cannot derive from experience. Nichols thinks that some beliefs—such as the belief that it is currently raining outside—may derive directly from raw experience—i.e., from experience not shaped by beliefs.¹ However, Nichols insists that indeterminism is too sophisticated a notion to grasp simply on the basis of experience: “. . . as an explanation for the belief in indeterminism, the appeal to experience is too anemic to be convincing” (2012: 293).²

Nichols suggests instead that indeterminist beliefs might derive from experience *shaped* by beliefs. After all, it is well known that background beliefs can alter perception in certain ways, via the phenomenon of cognitive penetration. Roughly, cognitive penetration occurs when the phenomenal character of one's experience is altered by one's cognitive states—for instance, by one's background beliefs or thoughts. If two subjects report different visual experiences from one another when looking at the same patch of color under the same conditions, the thesis of cognitive penetration says that these perceptual differences are to be ex-

¹ I explain what I mean by “shaped by beliefs” in the next paragraph.

² It is my aim in this chapter to challenge this claim.

plained by differences in the subjects' cognitive states, which in turn alter their respective experiences. There is ample evidence that this occurs in cases of visual experience (e.g., Delk and Fillenbaum 1965; Levin and Banaji 2006). If cognitive penetration occurs even in cases of visual experience, then in the absence of any reason to think that it cannot occur in cases of agentic experience, it is at least a plausible hypothesis that background beliefs may influence the character of experiences of freedom as well. As a result, even if experiences of being free to choose otherwise are initially compatibilist in character, that does not mean that they remain compatibilist. Incompatibilist beliefs might shape the experience and make it incompatibilist.

A problem with this explanation of indeterminist beliefs is that it is circular: it explains the relevant belief by appeal to experience, which is in turn shaped by indeterminist beliefs.³ As Nichols puts it, "If experience is supposed to provide a noncircular explanation for our belief in indeterminism (or our theoretical resistance to determinism), then it has to be in virtue of experience that is not guided by an indeterminist belief" (2012: 294).⁴

To avoid this problem, we might reconsider the possibility that Nichols rejects—namely, that indeterminist beliefs in fact have their source in experience.⁵ For instance, Richard Holton maintains that, "Our experience tells us that our choice is not determined by our beliefs and desires, or by any other psychological states—intentions, emotions etc.—to which we have access. Those could

³ Nichols' proposal that experiences of freedom are shaped by beliefs is also problematic since there is no positive evidence of cognitive penetration occurring in these experiences, however plausible this hypothesis might be.

⁴ The possibility remains that explaining the belief as deriving from experience is circular in an unproblematic way: it is a positive feedback loop. Experience by itself may be too "anemic" to be the source of indeterminist beliefs. But experience might influence beliefs in a minor way at first, and then beliefs might penetrate experience, and then experience in turn support belief, and so on, in a sort of "ratchet" or "bootstrapping" account. However, Nichols does not explore this possibility, and I leave it aside in what follows. Nichols thinks that "The idea of indeterminism ... is presumably much too complex to be directly given by raw experience" (294). It is this stronger claim that I wish to challenge.

⁵ This move is also motivated by the fact that there is no positive evidence that experiences of freedom are shaped by beliefs. See footnote 3.

be the same, and yet we could choose differently” (2006: 15). As a result of such experience, people form indeterminist beliefs.

Terry Horgan (e.g., forthcoming) agrees that people form indeterminist beliefs on the basis of experience. Yet he denies that the experience is indeterminist. As we saw in Chapter 3, Horgan concedes that people might *think* that they can tell just by introspection that their experience is indeterminist. However, this reflects a form of introspective error or confabulation. As Horgan puts matters, it is one thing to know (A) by introspection:

(A) My experience does *not* present my choice as determined by my prior states.

But it is another thing to know (B) by introspection:

(B) My experience presents my choice as *not* determined by my prior states.

Horgan agrees that it might be possible to know whether (A) is true by introspecting. Yet it is not possible to know whether (B) is true by introspecting. When one asserts (B), thus judging one’s experience as indeterministic, either one is mistakenly inferring (B) from (A) or conflating (A) and (B). Either way, the result is a fallaciously formed belief in indeterminism.

Nichols (2012) thinks that even this etiology for people’s belief in indeterminist freedom cannot be correct, since it fails as an adequate explanation of the relevant belief. We are not introspectively aware of the causes of our headaches either, Nichols argues, yet we do not infer indeterminism from that:

It would be a kind of scope fallacy to move from “I don’t experience my actions as determined” to “I experience my actions as not determined.” Now, people surely do commit fallacies. But notice we don’t seem to commit the scope fallacy when it comes to headaches. That is, the phenomenology of headaches doesn’t present us with a set of deterministic headache-causes, but we don’t leap to indeterminist conclusions there. (2012: 296)

Nichols proposes an alternative explanation for the source of belief in indeterminist freedom. First, he outlines how scientists judge whether a system is indeterministic. Researchers control for inputs to the system, and if it has different outputs given the same inputs, then one is warranted in concluding that the system is indeterministic (2012: 297). Next, Nichols presents an argument for indeterminism adapted from William of Ockham (2012: 298–99):

- (1) The factors that are introspectively accessible do not determine my choice.
- (2) I have introspective access to all the (proximal) factors that influence my choice.
- (3) Therefore, my choice is not determined.

Here, (2) is the crucial premise. One might, of course, think that this premise is false. Indeed, Nichols cites evidence showing that people do not assume that they have access to everything in their minds. However, (2) is actually a more limited claim: “All that is required is a kind of default (but defeasible) assumption that the causal influences *on decisions* [emphasis added] are introspectively available. And there is evidence that people do have such a default assumption” (2012: 299). So, people tend to believe (2) as stated. Furthermore, it seems that decisions are special in this regard. People do not have the same default assumption when it comes to whether all the proximal factors that influence their urges, for example, are introspectively accessible. Presumably, the same goes for headaches.

Why do people make the default assumption for decisions, but not for urges or headaches? Nichols thinks that it is because of a certain bias: “Our attention is drawn to factors that are present, rather than to the possibility that there are hidden factors” (2012: 301). Normally, Nichols maintains, people take themselves to have a good understanding of how mechanisms with discrete, accessible causal parts work—for instance, locks or zippers. (The relevant causal parts are “present.”) Yet when people are asked to explain how locks or zippers work, often they find themselves at a loss. As a result, people “downgrade” their level of

presumed understanding. By contrast, people do not think that they have a good understanding of how objects like flash drives work, given their lack of easy access to the causal parts of such devices. (The causal parts are “hidden.”) Nichols thinks that when people introspect on decisions, they find that they have access to discrete mental states (beliefs, thoughts, desires, etc.) that causally influence these decisions. As a result, they take decisions to be like locks or zippers. When they are presented with evidence from psychology detailing the actual, unconscious causes of their decisions, they may be brought to downgrade their level of understanding. That is, they may be brought to reject (2). People’s default setting is to think of decisions as like zippers, but of headaches as like flash drives. Yet just as people can be brought to downgrade their presumed understanding of zippers, they can also be brought to downgrade their understanding of decisions.

This explains why people tend to believe premise (2)—in other words, the claim that we have introspective access to all the proximal factors that influence our choice—as long as this claim about access is restricted to choices or decisions. It also shows that belief in indeterminism is not warranted, at least if it results from anything like an inference from (1) and (2) to (3). This is because people’s default assumption that they have introspective access to all the factors that influence their decisions is mistaken.

A major shortcoming of Nichols’ proposal is its backward-looking focus: it focuses on individuals’ access to the causes of their decisions. This is true of both Horgan’s and Holton’s accounts as well. By contrast, deliberation is centrally a forward-looking phenomenon. When agents look to the future as deliberators, two aspects of agency become especially salient, which Holton, Horgan, and Nichols all fail adequately to address: (i) the experience of having *alternative possibilities* for action, and (ii) the experience of being able to *decide between* such alternatives. These are some of the aspects of experience that libertarians most often cite as indeterministic.⁶

⁶ See e.g., Searle 1984: 95; C. A. Campbell 1951: 463; Ginet 1997: 89; O’Connor 1995: 196–197.

Against Nichols' claim, I maintain that people's indeterminist beliefs derive from their experience of navigating into the future, rather than from introspection on the causes of their decisions. I also maintain that such beliefs do not derive from a mistaken judgment about experience, as Horgan claims. Further, although such beliefs partly derive, as Holton thinks, from a feeling that one's choice is not determined, this is due in the first instance to the forward-looking character of deliberative experience. Nevertheless, while I concede that it is natural for agents to arrive at indeterminist beliefs on the basis of such experiences, these beliefs are not justified.

5.3 A Forward-Looking Model

A number of authors have recently embraced a forward-looking model of agentic experience.⁷ According to these authors, our experience as agents enables us—minimally—to distinguish self-generated actions from both involuntary bodily movements and the externally caused movements of objects. One prominent view of this sort is developed by Tim Bayne (e.g., 2011).

On Bayne's view, agentic experiences are produced by a dedicated perceptual system, which informs us about aspects of our own agency. This system includes forward models of action control. The forward models receive a copy of the agent's motor commands, which are used to predict the sensory consequences of the agent's bodily movements. The copy of the motor commands is also sent to a comparator, so-called because it compares the predicted sensory consequences of the agent's movements with sensory feedback.⁸ When the comparator identifies a match between prediction and feedback, it identifies the changes as self-

⁷ See, e.g., Bayne 2008, Synofzik et al. 2008, Bayne and Pacherie 2007, Blakemore and Frith 2003.

⁸ According to Bayne, the comparators lie between the standard perceptual systems and the motor system, since they take both perceptual representations and motor representations as inputs. In this way, the states generated by the comparators have the functional role of perception. That is to say, like other perceptions, the function of the *sense of agency* is to generate representation of some domain and make these available to the agent's cognitive systems in an experiential format (2011: 358–59).

generated. When there is no match (or a weak match), the changes are identified as externally caused.

In this way, the model explains how agents are able to distinguish—experientially—self-generated actions from (i) the externally caused movements of objects in the environment and (ii) their own involuntary bodily movements.

This is as far as Bayne extends his model. Yet one might think that it can explain more. In a suggestive passage on the experience of freedom, Horgan writes:

Some survival-important features of a creature's ambient environment will be ones that are susceptible to causal influence by suitable bodily motions by the creature itself, motions that can be internally generated by the creature's inner motion-control mechanisms. (Consider a bear, for instance. In an appropriately fortunate ambient environment in which there is a bush nearby with edible berries on it, there are potential bodily motions available to the bear that would have the effect of transferring some of those berries from the bush itself to the bear's stomach. For such potential bodily motions, the anticipatory-freedom phenomenology of "I can" (vis-à-vis those potential bodily motions) will be beneficial to the bear, as will the ongoing free-agency phenomenology experienced by the bear during feeding.) Other survival-important features of a creature's ambient environment will involve event-causal goings on that are *not* susceptible to causal influence by the creature's potential bodily motions, but that need attending to (and responding to) if the creature is to survive and flourish. (For instance, if a bear sees a huge boulder rolling down the mountainside in his direction, this ought to be registered by the bear as a state-causal sequence that (i) cannot be *influenced* by certain bodily motions, and (ii) will be big trouble if the bear's body remains where it currently is.) (Forthcoming)

The suggestion here is that something like Bayne's model might explain experi-

ences of having a general ability to do various things. Over time, an agent learns to identify, experientially, the kinds of options that exist for her as alternatives for action at a time, and these are picked out against a backdrop of phenomena that she experiences herself as powerless to affect (e.g., *eating those berries* vs. *stopping that boulder rolling down the mountainside*). This provides her with experiential representations both of alternative possibilities and of a general ability to do otherwise, and in such a way that each of these representations is uncontroversially compatible with determinism. The agent not only experiences (i) voluntary actions differently from externally caused movements, but also experientially identifies (ii) the kinds of alternatives that are amenable to her control at a time.

Even so, such a model is unable to explain libertarian experience, or to explain how indeterminist belief might derive from an indeterminist experience of choosing and deliberating. More would need to be said about how experiences of (ii) are generated. Bayne's model invokes only "systems that are concerned with motor control and production" (2008: 198), and thus it is not clear whether his model could possibly account for more than experiences of (i). In short, "the low-level contents of agentive self-awareness," such as the experience of self-generated actions or of externally-caused movements, might be "generated exclusively by low-level comparator systems" (198). By contrast, "the high-level contents of agentive self-awareness (such as the *kinds* [emphasis added] of actions that one takes oneself to be carrying out)," might have their etiology in another system, or in the comparator system operating together with some other system (198–99).

Nevertheless, Bayne thinks that we have a robust experience of freedom: "there is clearly an intuitive sense in which we can—and often do—experience ourselves as acting freely," although "[t]he difficulties that confront us in attempting to articulate what it is like to experience oneself as a free agent are . . . particularly imposing" (2008: 195–6). In particular, Bayne is unsure whether it is possible for anyone to experience libertarian freedom:

[P]erhaps experiential systems are incapable of inserting a negation quite where it needs to be inserted in order to represent libertar-

ian freedom. In order to represent an action as free in a libertarian sense one must not only represent it as undetermined by one's prior psychological properties but also as undetermined by one's physical properties—or indeed any physical properties. And it is not obvious that experiential systems have *that* kind of representational power. (2008: 196)⁹

Whatever the attractions of Bayne's comparator view might be, such a model cannot explain indeterminist experience, and it cannot explain either how anyone's belief in indeterminist freedom might derive from such agentic experience. The reason for this, I think, is that Bayne's model focuses on perceptual experience, and it is difficult to see how such experience could have as content that one is free to do otherwise, in any sense. To borrow Nichols' phrase, perceptual experience is too "anemic" to represent the possibilities required for such an experience. That would seem to require a comparison of two or more distinct representations—the alternatives themselves—in the mind, and that is clearly not a perceptual operation.¹⁰

In order to account for experiences of freedom, I turn to recent work done on prospection, which is the mental simulation of future possibilities. By embracing a forward-looking account of prospection, I will show how we can explain how people experience indeterminist freedom, and how they acquire a belief in such freedom. Even so, libertarian beliefs are not warranted.

⁹ This echoes the point made by Horgan and Nichols: even if we do *not* experience our choices as determined by our own prior states, it seems implausible that we also experience them as *not* determined by such states. Note also that Bayne's comment here seems at odds with his own view, since it slips into the backward-looking framework that considers only the causes of decisions, while his comparator model is importantly forward-looking.

¹⁰ As a consequence, it is not obvious how any development of Bayne's view could account for the sort of freedom at issue in free-will debates. Experiences both of general abilities and of the kinds of actions one can perform at a time are obviously compatibilist. No one thinks that general abilities are incompatibilist, and thus the *experience* of having general abilities could be accurate assuming determinism. So, even a development of Bayne's view along the lines suggested in the passage from Horgan would fall short of being an account of anything that is a matter of contention regarding freedom.

5.4 Etiology of Indeterminist Experience

Although Bayne is correct to focus on the forward-looking aspects of agency, his view does not go far enough. Bayne's view does not explain the sort of experience that might generate a belief in indeterminism. Still, unlike Nichols, I maintain that experience is exactly the right place to look for an explanation of people's belief in indeterminist freedom. Nonetheless, such experience is different from the sort on which Bayne focuses. The sort of experience that causes indeterminist beliefs results from the phenomenon of *prospection*,¹¹ which is the mental simulation of future possibilities for the purpose of guiding action. These simulations function as effective competitors to perceptual experience, so that they are actually experiential.

As I will show, the way in which *prospection* models choice and alternative possibilities for action makes it easy for deliberating agents both to experience and to believe that their freedom to do otherwise is indeterministic.

5.4.1 *Prospection is Forward-Looking*

Let us get a clearer picture of *prospection*. According to Martin Seligman, Peter Railton, Roy Baumeister, and Chandra Sripada (henceforth "SRBS"), "*Prospection ... is guidance ... by present, evaluative representations of possible future states. These representations can be understood minimally as 'If X, then Y' conditionals, and the process of *prospection* can be understood as the generation and evaluation of these conditionals*" (2013: 119).¹² Central to the phenomenon of *prospection* is its focus on how agents "navigate" into the future, as SRBS put it, rather than on how they are "driven by the past." Thus, *prospection* nicely captures the forward-looking aspect of deliberation and agency. According to this framework,

¹¹ The term '*prospection*' is due to Gilbert (2006), Gilbert and Wilson (2007), and Buckner and Carroll (2007).

¹² As we shall see, a useful way to think about the hypotheticals generated in *prospection* is in terms of causal modeling, as developed by Pearl (e.g., 2000) and Woodward (2003), among others.

...people and intelligent animals draw on experience to update a branching array of evaluative prospects that fan out before them. Action is then selected in light of their needs and goals. ... These prospects can include not only possibilities that have occurred before but also possibilities that have never occurred—and these new possibilities often play a decisive role in the selection of action. (2013: 119)

According to SRBS, in order to regulate an organism's interactions with its ambient environment, the brain must construct representational models of that environment.¹³ The most efficient models will be of the form, “*if* in circumstance C and state S, then behavior B has outcome O with probability p” (2013: 124). Like the comparator models of agentic experience developed by Bayne, these “feedforward/feedback” models take the following form:

(1) expectation → observation → discrepancy detection → discrepancy-reducing change in expectation → expectation → ...¹⁴

As SRBS put it,

Expectation is pivotal in schema (1) because it transforms experience into experimentation—continuously generating a “test probe” so that

¹³ SRBS argue that this is how a systems theorist would approach building an organism like us (Cf. Conant and Ashby 1970; Eykhoff 1994). They also note that this sort of approach is central to the prescient learning theory of Miller, Galanter, and Pribram (1960), as well as to both adaptive control theory (Åström and Murray 2008; Carver and Scheier 1990) and Bayesian epistemology (Earman 1992).

¹⁴ SRBS cite as empirical support for the phenomenon of prospection the near-optimal foraging behavior of various species (Dugatkin 2004). According to SRBS, “Foraging mammals have systems of neurons whose firing rates and sequences correlate with differences in: the identity of stimuli, their intensity, the magnitude of specific positive vs. negative hedonic rewards or food values, the relative value of a stimulus (e.g., deprivation vs. satiation), the absolute value of a stimulus (e.g., physiological need), the probability or expectation of a given outcome, the occurrence of a better- or worse-than-expected predicted error, and the absolute risk and expected value of given actions” (2013: 125; Cf. Craig 2009; Grabenhorst and Rolls 2011; Kringelbach and Berridge, 2009; Preuschoff et al. 2006; Quartz 2009; Rolls et al. 2008; Schultz 2002; Singer et al. 2009; Tobler et al. 2006). SRBS also cite recent evidence for the neural implementation of evaluative prospection during experiments with rats in T-mazes (Ainge et al. 2012; Derdikman and Moser 2010; Gupta et al. 2010).

the next experience always involves an implicit question and supplies an answer, which can then function as an error-reducing “learning signal.” (2013: 124)

In this way, agents generate and use mental simulations of future alternative possibilities, often by drawing on their past experience, for the purpose of enabling them to navigate into the future and select appropriate actions. The expectations generated by these simulations are tested against observed results in order to attenuate future expectations and actions. In this way, agents exercise “teleological” control over their decisions and actions:¹⁵

We call such accounts “teleological,” meaning explanation by selection in light of values and goals. . . . A good prospector must know more than the physical landscape—what is to be found where, with what probability—but also at what cost in effort and risk and with what possible gain. The prospecting organism must construct an *evaluative landscape* of possible acts and outcomes. The organism then acts *through* this evaluative representation, electing action in light of their prospects. And the success or failure of an act in living up to its prospect will lead not simply to satisfaction or frustration, but to maintaining or revising the evaluative representation that will guide the next act. To be sure, learning and memory necessarily reflect past experience. But at any given moment, an organism’s ability to improve its chances for survival and reproduction lies in the future, not the past. So learning and memory, too, should be designed for action. (2013: 120)¹⁶

¹⁵ SRBS allow that many sorts of everyday action do not require prospection, of either a conscious or an unconscious variety. When actions “can be successfully repeated without need for evaluation of alternatives” (2013: 125), they move from being under *teleological* control to being under *habitual* control.

¹⁶ While SRBS maintain that motivation is often teleological in the way that they outline in their “feedforward/feedback” model, SRBS concede that drive-like, non-goal-directed motivation *sometimes* occurs: “Addiction and salt deprivation, for example, can produce wanting without lik-

For SRBS, this forward-looking, desire-focused framework makes agency intelligible:

The driven-by-the-past framework makes agency and choice difficult to understand—individuals are responders rather than navigators. . . . If instead we see the individual as using past experience as information, as continually forming and evaluating a range of future possibilities, and as electing action from among these possibilities in light of what she likes and values, then we can see that active agency is a natural part of the causal structure of action. Motivation for such action is not determined by fixed drives or past conditioning, but is elicited by the evaluative process itself through the normal working of desire. (2013: 127)

In generating and using prospections in this way, agents are “drawn” toward the future by their evaluative representations, rather than “driven” by the past.

5.4.2 Prospecction Can Be Experienced

According to SRBS, prospecction typically occurs unconsciously and is unavailable to introspection, since it would be inefficient for agents consciously to keep track of all the prospections that they generate. Indeed, “even when individuals engage in conscious prospecction, their intuitive sense of the value of alternatives may be underwritten by unconscious simulation” (2013: 126; Cf. Railton in press). Yet prospecction can become conscious. One reason SRBS think it may be

ing . . . Certain physiological demands, natural or artificial, can produce ‘driven’ motivation even in the face of profound distaste and resistance, but this is atypical indeed. Ordinary action, even eating a meal when hungry, does not work this way—for hunger makes eating attractive, not distastefully compulsive” (2013: 127). Yet, normally, motivation depends crucially on desire, not on drive: “Philosophers since Aristotle have emphasized that desire is not a blind urge but rather represents its object as an ‘apparent good’ . . . or under a ‘desirability characterization’ . . . an attractive prospect that can elicit motivation to seek it—‘liking’ a representation gives rise to ‘wanting’ its object” (2013: 126). Cf. Aristotle ca. 330 BC/1999: 1113a15; Anscombe 1957: viii; Railton 2002; Berridge 2004.

useful for prospection to become conscious is that this enables agents to engage in *shared* prospection. Even though conscious prospection may be less efficient than unconscious prospection, conscious prospection might make for more effective prospection.

Whatever the benefits of conscious prospection might be, SRBS hypothesize that *affect* is central to how prospection becomes conscious, whenever it does. According to their story, when the process of prospection encounters “incommensurable dimensions and conflicting values and perspectives” (2013: 131), agents’ engagement in explicit, conscious comparison of these elements is facilitated by the brain’s “common metric” of affect:

Affect is the brain’s common currency for value, and conscious, subjective affect would permit the possible futures to be brought into the open for explicit comparison with each other. . . . conscious subjective affect attached to prospectations would enable them to compete effectively with ongoing experience. (2013: 131)

In other words, whenever agents have conflicting, incommensurable thoughts about what to do, their options feed into “an experientially rich and detailed workspace,” so that they can “use their intelligence and imagination to best effect.” In such cases, “it can be best to act in *awareness* [emphasis added] of . . . conflicting thoughts” (2013: 131). As a result, forward-looking prospection is consciously experienced by the agent. Nevertheless, this experience is not perceptual, in contrast to Bayne’s model.

5.4.3 Prospected Choices Are Free Variables

It is useful to think about the hypotheticals generated in prospection as carrying causal information about what *would* happen under certain variations in the values of exogenous variables in a causal model. A causal model is a representation that encodes hypothetical relationships between variables, where the variables

represent causal relations. According to causal modeling, when considering whether something is a cause, we ask, “What if things had been different?” and by answering this question we identify factors whose manipulation would produce changes in the outcome being explained. If this (cause) variable were altered in these ways, this (effect) variable would be altered in these ways. The main restriction on what counts as a variable is that it must represent particular events in such a way that they can be set to different values by interventions (Cf. Woodward 2003: 11–14). Thus, a variable that represents the event of my choosing dessert might take the value “1” if I choose cake, or “0” if I refrain from having dessert. An *intervention* in a model is an exogenous change to the value of a variable: we consider what happens in the model by tweaking just this variable’s value. (By contrast, an endogenous change to the value of a variable occurs because of the values taken by other variables within the model.) In this way, interventions are “surgical,” in the sense that the usual causes of a variable, or of a variable’s taking a given value, are ignored or suspended. When we causally model a situation, we “carve off” the situation from the rest of the world, and from its causal antecedents. We allow the variables whose antecedent causes we have thereby ignored to vary freely across a range of values, where this range of values has the following pragmatic restriction: none of the values should correspond to possibilities that we consider too remote (Hitchcock 2001: 286). The hypotheticals specifying the relations that hold between the variables in any causal model are stated as structural equations, which are asymmetrical in the following way: the values of the variables on the left hand side of the “=” are determined by the equations on the right hand side.

Assume, for instance, that you will choose dessert only if I choose dessert first or a friend joins us. Take the variable representing the event of my choosing dessert to be C , the variable representing the event of a friend’s joining us to be F , and the variable representing the event of your choosing dessert to be D . Of these, C and F are exogenous, while D is endogenous. Let C take the value “1” if I choose dessert; otherwise D takes the value “2.” Likewise, let F take the value “1” if a friend joins us; otherwise, F takes the value “2.” Finally, let D take the

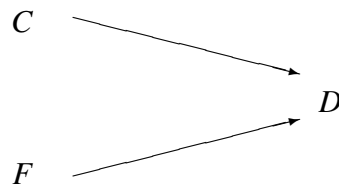
value “1” if $[C=1 \text{ or } F=1]$; otherwise, “2.” Now assume that I choose dessert and no one joins us. Thus, the structural equations that specify this extremely simple model are:

$$C = 1 \text{ or } 2$$

$$F = 1 \text{ or } 2$$

$$D = 1 \text{ if } [C=1 \text{ or } F=1]; \text{ else } 2$$

Here, the first two equations state the possible values that the exogenous two variables in the model, C and F , may take. The third equation encodes four hypothetical conditionals, two for each possible value of C and F . This equation says that you choose dessert only if I choose first or a friend joins us. These equations comprise the model, which may also take a graphical form indicating the dependency relations obtaining between the variables by means of “directed edges” connecting the variables in the graph:



Jenann Ismael (2013) thinks that agents mentally construct models of this sort when deliberating about what to do.¹⁷ Agents carve off the event of their making a choice from its causal antecedents, and treat it as an exogenous variable in a model, in order to assess the downstream effects of this variable’s varying freely across a range of values. In this way, agents capture causal information relevant to action-planning via the hypotheticals comprising the model. These are the same

¹⁷ The models I describe here are far simpler than those we presumably construct when deliberating about what to do. For one thing, my examples only involve variables with discrete values, whereas many models we construct in deliberating will have probabilistic values, as both SRBS and Ismael note. I have used simple models for the sake of clarity.

hypotheticals that SRBS claim our brain constructs in order to regulate interaction with the environment.

According to Ismael, a model, M , of a set of prospected options for any choice is “narrow-scope,” in the sense that it focuses only on a segment of the world—namely, the event of one’s making a choice and the prospected consequences of choosing in various different ways. In M , the event of one’s choice features as a free or exogenous variable, which we will call C . The variables that represent the prospected consequences are “downstream” of C , in the sense that the values they take are determined by the value C takes, together with the values of any other variables internal to the model (holding all other inputs and background conditions fixed). The downstream variables are endogenous and represent various possible consequences of one’s choice. The only way in which one can influence the values these variables take is by determining the value of the exogenous variable, C .

Imagine (once again) that C represents the event of my choosing dessert. Assume that C can take only two possible values: $C=1$ if I choose cake; $C=2$ if I decline dessert. Here, C is exogenous, and it is allowed to range across two values. Downstream of C , there is an endogenous variable, S , which represents the event of my falling asleep that night. Assume that the value S takes is determined solely by the value that C takes (holding all background conditions fixed). I know that if I choose cake it will (as always) keep me awake. If I decline dessert, I will fall asleep early. I have an important talk to give tomorrow morning, and I know I will perform at my best only if I go to sleep early. The event of my giving the talk is represented by L . Once C takes a value, the value L takes will be determined solely by S ’s value, and thus L is endogenous. The way in which C differs from S or L , in terms of how it functions within the model, is that C is allowed to range across more than one value in a way that does not depend on the values taken by any “upstream” variables. In this way, C is “carved off” from its antecedent causes.

In M , the set of prospected options is narrow-scope, given that it focuses only on a segment of the world: one’s choice. This contrasts with “wide-scope” mod-

els, which carve off larger chunks of the world. The narrow-scope model, *M*, might be embedded in a wider-scope model, *W*, in which the formerly exogenous variable *C* functions as a newly endogenous variable, the values of which are determined by the values of the exogenous variables of the wider model, together with the values of this model's other upstream endogenous variables (whose values would also ultimately depend on the values of the exogenous variables in the wide model).

To see how this works, consider a ball-and-socket joint in a robotic arm used in a factory assembly line. Recall that we select models on pragmatic grounds, according to the sort of question we are asking. If we are interested in how this joint works (perhaps we are engineers, designing a better joint), we will create a virtual separation of the joint from its environment and model it. The “frame” we put around this isolable causal structure has, at its boundaries, exogenous variables, which we allow to vary freely across a range of values. These variables' taking different combinations of values determines the values of the endogenous variables. When we insert the joint back into its environment—the robotic arm—we model the entire arm as a causal structure. As a result, the formerly exogenous variables that were allowed to vary freely become endogenous: the range of values that they can take is constrained by the exogenous variables at the boundaries of the more encompassing model. We will select this wider-scope model if we want to assess the efficacy of the robotic arm in an assembly line. Likewise, the event of an agent's choice might be embedded as an endogenous variable in a wider-scope model of social psychology.

The central point is that prospection treats the event of choice as exogenous, which effectively requires us to ignore its causal antecedents. This is important for two reasons. First, as I explain in the next subsection, it makes it natural for agents to experience (and also to think of) their prospected alternatives as ones that they can “get to.” Second, it shows how radically forward-looking deliberation actually is. In generating prospections, agents not only pay less attention to the causal antecedents of their decisions, and more attention to what SRBS call

a “branching array of evaluative prospects that fan out before them” (2013: 119), but mentally they *ignore* the antecedent causes. Nevertheless, prospected choices can additionally feature as endogenous variables in wider-scope models serving other purposes (like those of social psychology), where they feature as endogenous variables and thus have their causal antecedents “reattached.”

5.4.4 Indeterminist Experience and Judgment

Although SRBS weigh in briefly on the topic of free will, they do not address the sort of empirical evidence that I discussed earlier. This evidence indicates that people tend to experience possessing and to believe that they possess an indeterminist freedom to choose among alternatives. Even so, I claim that SRBS’s account of prospection, together with a causal-modeling account of how the hypotheticals generated in prospection should plausibly be modeled, can explain how people get to be indeterminists.

The experience of being free to do otherwise may seem indeterminist for two reasons. First, we experience an openness to the future that appears to require indeterminism for the experience to be accurate. Second, we experience choice as not being sufficiently caused by anything prior to it, and thus as not having antecedent deterministic causes. On the basis of such experiences, we tend to have indeterminist beliefs, which are subsequently revealed when we explicitly entertain the notion of determinism.

First, our prospected experience suggests an openness to the future that requires indeterminism to be accurate. Our experience when we engage in conscious prospection is that there is more than one way that we can extend the present into the future, depending on our choice. This makes it easy to experience (and to think of) our prospected alternatives as ones that we can “get to” in the sense that libertarians are trying to pick out, and which they claim might be provided for by indeterminism. After all, if determinism is true then there is only one physically possible extension of the present into the future. In that case, it would seem that our experience is inaccurate, and thus apparently the experience is incom-

patibilist. We need not be aware of this. Our experience might be, so to speak, *implicitly* incompatibilist: it might have a content, P, that is in fact incompatible with determinism. As a result, were we to entertain the thesis of determinism, we would *explicitly* judge our experience as incompatibilist.

In the language of causal modeling, we can put the point this way. Whenever we prospect future possibilities for action in the course of deliberating about what to do, the variable representing the event of our making a choice is exogenous, meaning that it is permitted to vary freely over a range of values. As a result, in our practical, deliberative experience we treat the event of making a choice as an exogenous variable in a narrow-scope model of deliberation. Yet, when we consider the same event while assuming determinism, we treat it as an endogenous variable in a wide-scope model. Here the variable is permitted to take just one value.¹⁸ This creates an apparent conflict between treating one and the same variable as exogenous and endogenous. This conflict makes the experience seem indeterminist.

Second, we experience our choice as not determined by anything prior to it. This is in direct denial of the claims made by Nichols, Horgan, and Bayne that even if we do *not* experience our choice as determined by our prior states, it is implausible that we experience it as *not* determined by such states. Why are they wrong about this? Recall that prospecting models choice as an exogenous variable, which is carved off from its antecedent causes and allowed to vary freely across a range of values. If we experience choice in this way, we experience it as *not* having antecedent sufficient causes. A choice that is experienced as not having antecedent sufficient causes is, a fortiori, experienced as not having antecedent deterministic causes.

If that is right, then we are able to explain people's belief in indeterminist freedom as being due to their agentive experience. People believe that the future is open in a way that requires indeterminism, since that is what they experience in prospecting. People also believe that their choice is not sufficiently caused by

¹⁸ Obviously, this value may be unknown, or even unknowable, prior to actually deciding.

anything prior to it, and thus implicitly it does not have antecedent deterministic causes. Of course, people's belief that their choice is not sufficiently caused by anything prior to it might be defeated once they learn that it *does* have such causes—perhaps even deterministic ones. Still, people's initial tendency will be to believe that it does not. Moreover, the experience itself will presumably remain unaltered.

The result of all this is that, contra Nichols' claim that experience is too "anemic" to be a plausible source of people's belief in indeterminism, indeterminist beliefs might well come from experience. Yet, contra Bayne's suggestion that indeterminist beliefs might come from perceptual agentive experience, such beliefs more plausibly come from the experience of prospection.

5.5 Belief in Indeterminist Freedom is Not Justified

Even if people have genuinely indeterminist experiences of freedom, it certainly does not follow that these experiences are accurate, or even that a belief in indeterminist freedom formed on the basis of such an experience is justified. First, even though it might easily seem that the future is open in a way that would require indeterminism for the experience to be accurate, all that is actually going on is that the agent is considering what outcomes she can cause, depending on what choices she makes, which requires letting the event of her choice range across more than one value. In this way, she can prospect the possible downstream effects of her choosing in different ways. Second, even if the agent introspects that her choice lacks deterministic causes, it does not follow that it actually lacks such causes, since the causes might not be introspectible. Third, the conditionals generated in prospection are subjective and relative, and so they cannot support the view that the future is indeterministically open. This is partly due to the epistemic nature of the possibilities that such conditionals capture, and epistemic possibility is obviously compatible with determinism.

Even so, that cannot be the whole story. We want to be able to say that agents

are free to choose among alternative possibilities—that is, that their freedom is a matter of fact, not something about their epistemic access to facts.

In closing, I want to suggest how experiences of indeterminist freedom are in fact compatible with determinism, in the sense of being accurate even if determinism is true. Further, I want to suggest that we are justified in believing that we are free to do otherwise, and that such freedom is consistent with determinism. This is because the claim that we are free to do otherwise should be assessed for accuracy according to the narrow-scope model that we use in deliberating about what to do. In that context, it is true. When the claim is instead assessed according to the wide-scope model of determinism, it is false. Even so, there is no conflict between these models.

Recall: we select models on pragmatic grounds. In a situation of choice, the model that we naturally select treats the event of our making a choice as an exogenous variable. This occurs automatically. The apparent threat from determinism comes from the thought that wide-scope models in some way override narrow-scope models. As Ismael puts it, “To get the purported conflict with free will going we are invited to see action in the context of wider embedding models . . . The worry about physical determinism is the most extreme version of this sort of model” (2013: 229–230).

Ismael argues that the rules governing narrow-scope models do not derive from the rules governing the wider-scope models in which they are embedded. In fact it is the other way round: we “start with the basic building blocks with a great deal of freedom of movement and build up more complex systems by restricting their relative motion” (2013: 227). That is to say, the rules governing the whole are derived from the rules governing the parts, not vice versa. Further, the rules governing the parts provide agents with richer causal information than the rules governing the whole. This is because “[v]ariables that were allowed to vary freely in the original model are constrained by the values of variables in the embedding model and so we just lose information about what *would* happen *if* they were

allowed to vary freely” (2013: 218). As a result, the wide-scope model of determinism does not override the narrow-scope model used in prospection.

The correct model to adopt when deliberating about what to do is the narrow-scope model, since it is more useful relative to the pragmatic purposes of making a choice. This model is compatible with wider-scope models. We develop wider models by taking narrower models and adding constraints on the values that their exogenous variables can take. Thus, we arrive at the maximally wide-scope model of determinism by taking models of things like prospected choices and restricting the value that each variable in such a model can take to just one. However, as Ismael puts it, “there is no more conflict between these models than there is between the view of a building from close-up and the view from a very great distance” (2013: 230).

Ismael thinks that her naturalistic picture of causation and deliberation is “faithful to the experience of agency” (2013: 232). She is right about this in more ways than she explores. Even if people have indeterminist experiences of freedom (a possibility that Ismael does not consider), these experiences are compatible with determinism, in the sense of being accurate even if determinism is true. This is because the accuracy of such an experience in a given context of deliberation is to be judged according to the narrow-scope model appropriate to that context. Moreover, if people form a belief in indeterminism on the basis of such an experience, not only is this belief not justified, but the opposing belief may be justified—namely, the belief that freedom is *consistent* with determinism. In the next two subsections, I sketch in more detail how this proposal is meant to work.

5.5.1 Radical Compatibilism

Consider the feeling that an agent, S, might have about her openness to the future and her freedom to do otherwise: “It feels like I can A at t , or refrain from A-ing at t by B-ing instead.” In order for this experience to be accurate, the following would have to be true for S: (ATDO) “I can A at t , or refrain from A-ing at t by B-

ing instead.”¹⁹ On the usual way of assessing the truth of such a claim, traditional compatibilists insist that we may consider possible worlds with differences in the events antecedent to *t* or in the laws of nature. Incompatibilists insist that we should only consider worlds with the same antecedent events and laws as the actual world.

On the present way of looking at matters, however, the relevant hypotheticals are not assessed according to the usual semantics for counterfactual conditionals.²⁰ Rather, they are assessed according to the structural equations comprising the relevant causal models. In a deliberative context, (ATDO) may be true when assessed according to the model relevant to that context, i.e., a narrow-scope model in which the event of choice features as an exogenous variable. Yet when (ATDO) is assessed according to the wide-scope model of determinism, it is false. In such a model, there is just one value that the variable representing the event of choosing can take. When trying to decide whether to A, that is not a useful model to adopt.

Imagine that S judges (ATDO) as false when she explicitly assumes the truth of determinism, defined as the thesis that there is only one physically possible future, given the past and the laws. So S judges her experience of being free to do otherwise as incompatibilist (i.e., as inaccurate if determinism is true). Thus, presumably her experience had a content that, when considered in relation to determinism, is judged as inconsistent with that thesis.

Perhaps a traditional compatibilist will tell us that S is making a mistake in judging that (ATDO)—and so her experience of being free to do otherwise—is false under the assumption of determinism. Somehow, S is assessing the truth of (ATDO) by considering worlds with the same antecedent events and laws as the actual world, and this is not what she *should* do. Yet, if there is something in the content of S’s experience of being free to do otherwise that encourages her to

¹⁹ Here, “ATDO” stands for “Ability To Do Otherwise.”

²⁰ Problems have been suggested for possible-worlds analyses of counterfactuals (e.g., Barker 2011, Fine 2012).

do this, then the traditional compatibilist's claim will be idle: S's experience is inaccurate, and that is that.

On the present proposal, however, nothing about the content of S's experience is altered, yet that very experience is accurate, even assuming determinism. This is in stark contrast to the traditional compatibilist's proposal. Let us call the present view *radical* compatibilism. By all means, says the radical compatibilist, (ATDO) is false when assessed according to the structural equations that encode the hypotheticals comprising the maximally wide-scope model of physical determinism. (Traditional compatibilists will not concede this.) Nevertheless, this verdict is consistent with judging (ATDO) as true when assessed according to the narrow-scope model of S's situation of choice. So, the truth of (ATDO) is consistent with determinism, even when there is something implicit in the content of a person's experience that encourages her to judge it as incompatibilist.

If one's belief about being free to do otherwise derives from this sort of experience, then what amounts to an *implicitly* libertarian belief about freedom turns out to be consistent with determinism. Only when one makes an *explicit* incompatibility judgment is one's belief false. What might appear to be a libertarian belief is not incompatibilist, when properly understood.

5.5.2 Contextualism Revisited

The view described in the previous subsection is a contextualist proposal: in a deliberative context, the claim that the agent is free to do otherwise is true, since it is assessed according to the narrow-scope model of a situation of choice; but this very claim is false in a context in which it is assessed according to the maximally wide-scope model of physical determinism. Recall that in Chapter 3, I criticized Horgan's contextualist proposal. Thus, it may be thought that there is a *prima facie* tension between my endorsing contextualism here, yet my criticizing it there. However, this is not so.

First, notice that the two forms of contextualism are distinct. On the sort of contextualist view endorsed by Kratzer (1977), which I sketched in Chapter 1,

'can' is treated as an existential quantifier over possible worlds restricted by an "in-view-of" clause. Thus, for instance, it may be true that "In view of the rules of chess, you cannot move the pawn three spaces ahead," while it may be false that "In view of your physical constitution, you cannot move the pawn three spaces ahead." Likewise, as we saw in chapters 1 and 3, Lewis (1986) gives a treatment of 'can' in terms of restricted possibility: to say that 'S can A' is to say that S's A-ing is compossible with certain facts, where the relevant facts depend on the stringency with which 'can' is used. This is the sort of view that Horgan adopts when he maintains that claims about S's freedom to do otherwise are governed by implicit, contextually variable, semantic parameters. On this view, simply asking the compatibility question about freedom and determinism drives the semantic parameters far beyond their normal settings to a maximally strict setting, according to which any claim that S is free to do otherwise is false.²¹ However, the falsity of such a claim is established according to a restricted possibility metric, such as Lewis's or Kratzer's, which uses a possible-worlds apparatus to assess the truth of the claim. This is not how the sort of contextualism I outlined in the previous subsection works. Although we arrive at the wide-scope model of determinism by restricting the possibilities available in narrower-scope models, this restriction is not implemented by means of quantification over possible worlds. Rather, it is implemented by embedding narrow-scope causal models in wider-scope models, and thus limiting the values that the previously exogenous variables in the narrow-scope model can take, once they become endogenous variables in a wider-scope model. Moreover, the truth of the claim that S is free to do otherwise issues from the model (and the hypotheticals that it encodes), which is selected on pragmatic grounds. The claim is not assessed according to any standard possible-worlds semantics.

Further, I argued in Chapter 3 that it was a problem for Horgan's contextualism that worries about whether we are free to do otherwise arise even when contextual

²¹ It is again worth noting that the traditional compatibilist view sketched in the previous subsection will not grant this.

parameters are normal. If that is right, then Horgan's claim that such worries arise only when we raise the parameters beyond their normal settings and explicitly ask the compatibility question is false. Even when we apply ordinary standards and do not explicitly invoke determinism, it seems we can generate worries about whether people are free, even in the sense of being able to do otherwise. Thus, it is not clear whether any scorekeeping confusion occurs when we raise the parameters and explicitly ask the compatibility question about experiences of being free to do otherwise and determinism. We may simply be exhibiting our competence in applying the notion of freedom in that context as well.

In Chapter 3, I demonstrated this point by considering Frankfurt-cases, in which Black—a neurosurgeon—wants Jones to choose A. Black can intervene to control Jones's brain processes should Jones be about to choose B. Yet Black prefers not to intervene unless he has to. Instead, he waits to see how Jones will choose on his own. Jones is unaware of Black's presence. Frankfurt claims that Jones lacks alternative possibilities in the case. Nevertheless, if Jones chooses A on his own, then apparently he freely makes his choice even though he has no alternative. Despite this, it seems that we can ask whether it is reasonable to expect that Jones have done something else instead, given that the conditions in which he found himself ruled out any alternative. If we think it reasonable to expect that Jones not have A-ed (where A-ing is killing Smith), then we have located a conflict in our thinking about how to apply the notion of freedom. If we consider the case just by focusing on the intervener, without considering determinism, we might want to grant—given that Black did not intervene—that Jones freely killed Smith. After all, he killed Smith on his own. Yet it is not clear whether it is reasonable to expect that Jones have done something else. Recall, he was unable to do otherwise. Did Jones freely kill Smith? Perhaps not. Have we illicitly raised the contextual parameters governing application of the relevant notion of freedom? It is not clear that we have. Once we point out that determinism is meant to function in the same way as Black in the case, by blocking the availability of Jones's alternative possibilities and thus blocking his ability to do otherwise, we

have generated a worry about Jones's freedom according to the ordinary standards governing application of this notion.

Notice that this way of understanding Frankfurt-cases relies on the idea that Black functions in the same way as determinism in the case. As I pointed out in Chapter 3, libertarians might deny this claim. They might insist that Black does *not* block alternatives, even though they will grant that determinism *does* block alternatives.²² However, I noted in Chapter 3 that even if Black does not block all alternatives, he plausibly blocks the sorts of alternatives that would be required for the ability to do otherwise. Conversely, I noted that a compatibilist about modal freedom might insist that determinism *does* allow for a compatibilist ability to do otherwise, even if Black does *not*, and so Black does not function in the same way as determinism. The problem is that most compatibilists actually grant that determinism does block alternatives, and thus blocks the ability to do otherwise.

What I want to show is that the causal modeling framework on which the contextualist view that I dubbed "radical compatibilism" depends enables us to show that Jones *has* relevant alternatives in a Frankfurt-case, and therefore *is* free to do otherwise. So: Black functions differently from determinism in the case. As a result, when the claim that Jones is free to do otherwise is assessed according to the hypotheticals comprising the relevant model (which is the relevant context of evaluation for the case), it is true that Jones is free to do otherwise, even though when the same claim is assessed according to the maximally wide-scope model of determinism, it is false. The upshot is that a Frankfurt-case does not generate any worry about modal freedom when judged according to the relevant causal model, whereas on Horgan's contextualist proposal, it *does* generate such a worry. Thus,

²² The idea here is that if the prior sign by which Black knows that Jones is about to B (e.g., a neurological pattern) is a *deterministic* predictor of Jones's A-ing, then it begs the question against the libertarian by assuming determinism in the case. Yet if it is an *indeterministic* predictor, then Black does not know for sure that Jones will A, and so Jones retains alternative possibilities in the case.

the form of contextualism that I endorse is free from the criticism that I levelled at Horgan's brand of contextualism in Chapter 3.

How does this work? In a Frankfurt-case, we assume that Black has access to a prior sign ($PS=2$) that predicts whether Jones is about to decide to B ($JD=2$). If so, then Black intervenes ($BL=1$) to ensure that Jones decides to A ($JD=1$), such that Jones A-s ($JO=1$). Otherwise, $PS=1$, Jones decides to A on his own ($JD=1$), Black sits idly by ($BL=2$), and Jones A-s ($JO=1$). Thus, the model looks like this (variables, equations, graph):

Variables

$PS=1$ if prior sign occurs that Jones is about to decide to A; $PS=2$ if prior sign occurs that Jones is about to decide to B

$BL=1$ if Black intervenes to ensure that Jones decides to A; $BL=2$ if Black sits idly by without intervening

$JD=1$ if Jones decides to A; $JD=2$ if Jones decides to B

$JO=1$ if Jones does A; $JO=2$ if Jones does B

Equations

$PS=1$ or 2

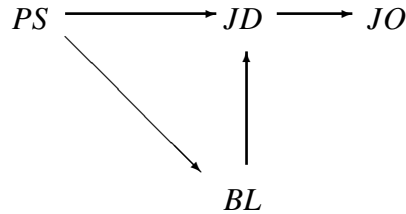
$BL=1$ if $PS=2$; else 2

$JD=1$ if ($PS=1$ or $BL=1$); else 2

$JO=1$ if $JD=1$; else 2

Actual Case: $PS=1$; $BL=2$; $JD=1$; $JO=1$

The corresponding graph is:



In the actual case, $PS=1$ (rather than $PS=2$) is an actual cause of $JO=1$ (rather than $JO=2$).²³ Yet if (counterfactually) $PS=2$, then it would be true that $BL=1$, $JD=1$, and $JO=1$.

In the causal model, we intervene on the value of Jones’s decision (JD) in order to see what happens, while holding fixed other direct causes of JO . That is how we test whether a variable is a difference-maker. In this case, there are no other direct causes of JO , so when we intervene on JD we not only ignore PS but also BL as a direct cause of JD . Ignoring or suspending both these variables as inputs to JD is necessary in order to assess whether $JD=1$ is an actual cause of $JO=1$, i.e., in order to assess whether Jones’s decision considered on its own (while ignoring Black) is the relevant causal variable in the situation. Note that here JD is allowed to range freely over two values, which represent Jones’s deciding either to A or to B, and so the causal model allows us a way of saying that Jones’s doing what he does rather than doing something else is the difference-maker in what happens in a Frankfurt-case. This is because when we focus on Jones’s decision, we screen off or ignore the variable representing Black’s possible intervention. When we do so, notice that Jones is (in a relevant sense) free to do otherwise. What the modeling does is illustrate our intuition that Jones is the difference-maker in the actual case, even though there is another potential cause that would ensure the same outcome. More importantly for present purposes, since the variable representing Jones choice (JD) ranges over two values, he is free to do otherwise.

²³ I am explicitly combining the interventionist model with a contrastive account of causation, as I suggest in Deery (forthcoming). The model I adopt is from co-authored work with Eddy Nahmias (unpublished).

This way of understanding a Frankfurt-case reveals that Black does *not* function in the same way as determinism. This is because Black's choice to intervene is a variable in the model rather than (as determinism would be) part of the background conditions that are being held fixed. Thus, Frankfurt-cases do not generate global worries about freedom when judged according to the relevant causal model, whereas on the sort of contextualism that Horgan endorses, they *do* generate such worries. As a result, the form of contextualism I endorse in the present chapter is free from the sort of criticism that I made of Horgan's contextualist proposal in Chapter 3.

5.6 Conclusion

The view that I have outlined accomplishes two things. First, it explains why some people believe that being free to do otherwise is incompatible with determinism: they do so because of their experience as deliberating agents, due to the phenomenon of prospectation. Second, it explains how these apparently incompatibilist beliefs and experiences are compatibilist. The experience of having alternative pathways into the future, and so of being free to do otherwise, is compatible with determinism. Furthermore, any implicitly incompatibilist belief that has its source in such an experience is also compatibilist, when considered correctly. That is to say, even though the claim that one is free to do otherwise is correctly judged as false when assessed according to the maximally wide-scope model of determinism, this is consistent in a relevant way with the judgment that the claim is true when assessed according to a model appropriate to a deliberative context. This type of contextualism is free from the problem that undermines Horgan's contextualist proposal, namely, that his proposal generates global worries about freedom even in when the context is ordinary.

Of course, the view described in this chapter falls short of providing a comprehensive compatibilist theory of the freedom to do otherwise. That remains as a task for another day. It remains open what properties and capacities such freedom consists in. Nonetheless, on the view that I have outlined here, our beliefs

and experience regarding such freedom—whatever it ends up actually consisting in—are perfectly compatible with determinism.

Chapter 6

Conclusion

In this chapter, I review the dissertation and show how its various strands hang together.

6.1 What the Dissertation Achieves

This dissertation addresses the question whether experiences of freedom are consistent in a certain way with determinism. The thesis of determinism says that a statement of the non-relational facts of the world at a time, together with a statement of the laws of nature, entails all other facts about the world at other times. So, determinism says that, at any given time, there is exactly one physically possible future. Throughout the dissertation, I take ‘free will’ to mean the freedom to choose among alternative possibilities for action. In doing so, I rely on a distinction between *moral freedom* and *modal freedom*. We have moral freedom when we control our actions in the strongest manner necessary for being morally responsible for them, whereas we have modal freedom when we are free (or able, in the relevant sense) to perform a given action at a time, or to refrain from performing that action at that time. In the dissertation, I focus exclusively on modal freedom (i.e., the ability to do otherwise), while leaving it entirely open whether such freedom is required for moral freedom.

I take modal freedom to have two aspects: (a) an alternative possibility aspect, and (b) an ability aspect. The alternative possibility aspect is necessary, but not sufficient, for modal freedom. The ability aspect involves the power of an agent to originate changes in the environment, or to “make things happen.” This is an intentional, not merely a simple ability. You have a simple ability to roll a 6 on a die as long as (for instance) your arm is working sufficiently well to roll a die. Yet a simple ability is not an intentional ability: you cannot ever intentionally roll a 6. I am interested in intentional ability. An account of intentional ability will involve simple ability, plus a story about intentional action, where for an event to be an intentional action requires that it be non-deviantly caused by certain of one’s mental states (such as belief and desires). So, modal freedom is, first of all, an intentional ability: it is the ability to perform intentional actions.

Modal freedom is also a specific, not just a general ability. I am generally able to do many things that I cannot do right now (even things that normally I can intentionally do), because I lack the opportunity to do them. For example, I cannot play the guitar right now, even though I am well able to play the guitar. This is because I do not have a guitar to hand now.

In Chapter 1, I characterize specific ability as follows. An agent has a specific ability to perform a given action at a time only if (i) she has a general ability to perform it at that time; (ii) she has an opportunity to perform it (which, minimally, requires that she have a certain possibility for action available at the time); and (iii) holding fixed the past and the laws (including the agent’s motivations regarding her opportunities as they are), the agent can exercise her general ability to perform that action at that time. If an agent has a specific ability in this sense regarding two distinct actions, A and B, at a time, then she has the specific ability to do otherwise. That is modal freedom: it is the specific ability to intentionally perform an action at a time, or to refrain from performing that action at that time.

My focus is on the compatibility question about modal freedom: Is such freedom compatible with determinism? To begin with, though, I focus on the “natural” compatibility question, which is a descriptive question about people’s pre-

theoretic belief-tendencies about their own and others' agency. Here, I rely on a distinction between three sorts of question about free will: descriptive, substantive, and prescriptive. The descriptive question is the question that I have just outlined: it asks about people's pre-theoretic belief-tendencies. The substantive question asks whether people are in fact agents of the sort that they tend to believe they are, in a given respect. Finally, the prescriptive question asks how best to theorize about freedom, given how we have answered the descriptive or substantive questions.

There are two ways to address the descriptive question: (a) ask people for their intuitions or beliefs about modal freedom (a good deal of experimental philosophy does exactly this), or (b) ask people about their *experiences* of modal freedom. I focus on the second method, concerning people's experiences of freedom. One advantage of this approach is that it is plausible that experiences of freedom partly drive people's beliefs (or intuitions) about freedom. However, focusing on experiences requires slightly reframing the compatibility question, and redefining compatibilism and incompatibilism. Those whom I call *experience-compatibilists* about modal freedom think that experiences of such freedom are sometimes veridical (or accurate) even if determinism is true, whereas *experience-incompatibilists* think that such experiences are non-veridical (or inaccurate) if determinism is true.

In the dissertation, I present new evidence supporting experience-incompatibilism on the descriptive question: people's experiences (and thus beliefs) tend to be incompatibilist about modal freedom. By contrast, I leave the substantive aside as a project for another day. Finally, in chapters 3–5, I defend compatibilism on the prescriptive question: we *should* be compatibilists, I argue, about experiences of (and beliefs about) modal freedom.

6.2 New Evidence on the Descriptive Question

In Chapter 2, I report the results of studies I conducted with Matt Bedke and Shaun Nichols (Deery et al. 2013). These results support experience-incompatibilism on the descriptive question about experiences of modal freedom.

In the experiments that my colleagues and I ran, we found that participants described their experience of modal freedom as inconsistent with determinism whether the decision was (i) present-focused or retrospective, (ii) imagined or actual, or (iii) morally salient or morally neutral. The only case in which participants did not report incompatibilist experience was when the question was explicitly about whether experiencing ignorance of the future is compatible with determinism—i.e., when the question was about *epistemic* openness, which is obviously compatible with determinism.

These results support experience-incompatibilism on the descriptive question. They also show that existing compatibilist accounts of freedom are inadequate to people's reported experience. So, if we want to be compatibilists, it turns out that we have got work to do on the prescriptive question.

In chapters 3–5, I turn to the prescriptive question. I begin by assuming (at least for the sake of argument) that there are experiences of modal freedom with libertarian content. First, I defend this hypothetical claim against an error theory for incompatibilist reports about the experience of freedom, and then I argue that even assuming that there are such experiences, they do not threaten experience-compatibilism. Finally, in Chapter 5, I develop an etiological story about the source of libertarian experiences.

6.3 Against a Compatibilist Error Theory

Chapter 2 describes results that support experience-incompatibilism on the descriptive question. But do they support incompatibilism? In Chapter 3, I consider an error theory for incompatibilist judgments about experiences of freedom—a view that says such judgments are somehow mistaken.

For instance, Terry Horgan (e.g., 2012; 2011) agrees that people often think that their experience is incompatibilist. Yet he argues that even when people *judge* their experience as incompatibilist, actually it is compatibilist: people misinterpret their experience. By spelling out how this happens, Horgan develops an error theory for incompatibilist judgments about experience. If Horgan's error theory is right, then it threatens to undermine the results of the experiments that I report in Chapter 2.

First, Horgan thinks that when we pay attention to certain aspects of our current experience, there is no significant appearance-reality gap. However, when we make a *judgment* about what we are paying attention to, it turns out that we make mistakes, especially when we try to make sophisticated judgments like judging whether our agentic experience is compatible with a general hypothesis about the world, such as the thesis of determinism.

Even so, Horgan concedes that many people (including many philosophers) actually judge their experience of freedom as incompatibilist. Horgan must explain why people make such judgments. He does so in two ways. First, he suggests a way in which people introspectively confabulate. Second, he tells a contextualist story about the application conditions of the notion of freedom, which also applies to judgments about experiences of freedom.

In this review, I focus on the first explanation. Horgan suggests that if we think that we can tell by introspection that our experience is incompatibilist, we are mistaken. It is one thing, Horgan thinks, for us to claim to know the following by introspection: (A) My experience does *not* present my behavior as determined by my prior states. But it is another thing entirely to claim to know the following just by introspecting on phenomenology: (B) My experience presents my behavior as *not* determined by my prior states. The first claim is unproblematic. Yet we cannot ascertain the truth of the second claim by introspection alone. When we make the second claim, i.e., when we judge that our experience presents our behavior as *not* determined by our prior states—and thereby judge our experience as

incompatibilist—either we are mistakenly inferring this claim from the innocuous first claim, or else we are conflating the two claims.

Even if we grant Horgan’s hypothesis about introspective confabulation, I maintain that more needs to be said about how such a mistake occurs. Notice that people do not make this sort of mistake when it comes to headaches. We do not mistakenly infer from the claim that our experience does *not* present our headache as determined the further claim that we experience our headache as *not* determined. Thus, Horgan needs to say how the experience of deliberation is relevantly different from that of headaches. This requirement is a theoretical cost of Horgan’s view, which the alternative position that I develop in chapters 3 and 4 does not incur. Indeed, while Horgan’s view is attractive in many respects, it incurs a number of theoretical costs that do not need to be incurred by compatibilists.

A central theoretical cost of Horgan’s view is that it requires maintaining that the content of experiences of freedom is exhaustively compatibilist. This claim enters compatibilists into intractable disputes with libertarians about the content of experiences of freedom. By contrast, it is a theoretical advantage of my view that it does not claim that experiences of freedom have exhaustively compatibilist content. This enables me to avoid entering into intractable disputes with libertarians. On my view, even if we take people’s incompatibilist reports about their experience of freedom at face value, and thereby grant both that *introspection reliably latches onto the content of such experience* and that *such content is rich enough to be incompatibilist*, there is still an important respect in which the experience is compatibilist: it is veridical even if determinism is true. I sketch this view at the end of Chapter 3 and develop it at length in Chapter 4.

6.4 An Alternative Compatibilist Proposal

On my view, experiences of freedom have two sorts of phenomenal content. If one type of content is libertarian, yet an experience of freedom also has another sort of content that is compatibilist, the experience might be veridical if the latter

content is satisfied. David Chalmers (2006) makes a similar move in connection with the phenomenal content of color experience.

Chalmers thinks that the view about phenomenal content that is most adequate to the phenomenology of visual color experiences is *primitivism*. According to the primitivist view, phenomenology presents colors to us as simple intrinsic properties of objects, spread out over their surfaces. As a result, experiences of color have contents that attribute primitive properties. A problem with this view is that there is good reason to think that the relevant primitive properties are not instantiated in our world. Thus, according to primitivism, none of our experiences of color is veridical.

Chalmers argues that there is also another type of phenomenal content that makes color experiences accurate, at least in the right kinds of cases. This is ordinary content, which has its own veridicality condition: that the relevant object have whatever property (or set of properties) normally causes phenomenally red (etc.) experiences. Here, the phenomenal content is a “condition on extension.” However, there is also a problem for this view: ordinary content is not adequate to the phenomenology of color experience, since it does not reflect the phenomenal character of such experience.

Chalmers’s idea is to combine these views in a way that captures the truth-conditional virtues of ordinary content and the phenomenological virtues of primitivist content. I do something similar for experiences of freedom.

I begin by assuming (at least for argument’s sake) that the libertarian content of an experience of modal freedom is of one’s feeling a certain unconditional openness to the future. The future feels open in a way that would require indeterminism for the feeling to be accurate. It feels “as if” one is free to decide to A or, in an unconditional (i.e., libertarian) sense, to refrain from A-ing. Of course, it may seem that an experience with such libertarian phenomenal content is veridical only if libertarianism is true. However, my view shows that experiences of freedom with libertarian content might still be consistent with determinism, in the

sense of being veridical even if determinism is true (and thus, even if libertarianism is false).

My proposal (by analogy with Chalmers's move) is that there is a second sort of phenomenal content to experiences of freedom that is compatibilist. This content is a condition that a property must satisfy in order to *be* the property that is attributed by the experience. The property attributed by the experience is, of course, the freedom to do otherwise. What condition might work as the second phenomenal content for such an experience?

For color, the second phenomenal content Chalmers's proposes is the following condition: whatever property (or set of properties) ordinarily causes phenomenally red (etc.) experiences. In the agentic case, I think we cannot say that the second phenomenal content is whatever property (or set of properties) ordinarily causes experiences of being free to do otherwise, since presumably no one thinks that the meaning of 'being free to do otherwise' is *being such that it causes an experience of being free to do otherwise*. More plausibly, the second content is the following condition: That there is instantiated whatever relevant property (or set of properties) is ordinarily instantiated when one experiences being free to do otherwise.

There is good reason to think that this condition picks out a genuine second phenomenal content for experiences of freedom. In the dissertation, I explain the plausibility of this content by analogy with a natural-kind view about modal freedom. This view says that 'freedom' (or 'free act') is a natural-kind term that refers to whatever relevant processes are at work in choices or decisions that we ordinarily call 'free,' or that feel free to us, whatever those processes turn out to be (as long as they constitute a relevant kind).

I also explain how libertarian and compatibilist phenomenal content are related in the following ways. For an experience of being free to do otherwise to be *perfectly* veridical, we would have to live in a world where libertarianism is true. The best that we can do if determinism is true, by contrast, is to have certain properties "match" the libertarian properties that are attributed by the presenta-

tional content, by playing the role that libertarian properties would play in such a world. So, the libertarian phenomenal content sets an ideal standard for veridicality, and the second, compatibilist content is a condition that relates us to whatever properties come closest to meeting this ideal standard. Once the second content is satisfied, the experience is *imperfectly* veridical, even if determinism is true.

Thus, despite granting that our experience of freedom has libertarian content, this experience might be veridical under the assumption of determinism.

6.5 The Source of Libertarian Content

In Chapter 5, I address the question where libertarian content might come from. Libertarians often think that people's experiences of deliberating and choosing have libertarian content, and that such experiences lead people to believe that their choice is indeterministic. The studies that I report in Chapter 2 indicate that libertarians may be right about this. However, Shaun Nichols (2012) argues that belief in libertarian freedom cannot have its source in experience. Nichols argues that raw experience is too "anemic" to have content that requires indeterminism to be true. Nichols instead proposes an alternative etiology for belief in libertarian freedom.

I argue that a major shortcoming of Nichols' view is its backward-looking focus: it relies on the idea that people's indeterminist beliefs derive from introspection on the causes of decisions. Yet, deliberation and choice are crucially forward-looking. When agents look to the future in the course of their deliberating about what to do, two aspects of agency become especially salient, which Nichols does not adequately address: (i) the experience of having *alternative possibilities* for action, and (ii) the experience of being *free to choose between* such alternatives. These are some of the aspects of experience that libertarians most often cite as indeterministic.

By contrast, Tim Bayne (2011) embraces a forward-looking model of agentic experience. Bayne thinks that agentic experience enables us to distinguish our self-generated actions from both involuntary bodily movements and the exter-

nally caused movements of objects in the environment. Yet Bayne's view cannot explain people's experience of indeterminist freedom, and so it cannot explain how people's belief in libertarian freedom derives from agentic experience. This is because Bayne's view focuses on perceptual experience, and it is difficult to see how perceptual experience could have as content that one is free to do otherwise. That would require a comparison of two or more distinct representations—the alternative possibilities themselves—in the mind, and that is not a perceptual operation.

I explain people's experience of libertarian freedom, and thereby the source of people's belief that they possess such freedom, by appeal to the phenomenon of prospection, which is the mental simulation of future possibilities for the purpose of guiding action (Cf. Seligman et al. 2013). Crucially, prospection can be experienced, and because of the way in which the hypotheticals generated in prospection should be modeled in an interventionist framework, it turns out to be easy for deliberating agents both to experience their choice as indeterministic, and to believe that their choice is indeterministic (and thus libertarian), even though the modeling itself is consistent with determinism. Prospection treats the event of one's making a choice as an exogenous variable in a model of prospected outcomes, and (as I outline in Chapter 5) this requires ignoring its causal antecedents. This way of modeling choices is important for two reasons.

First, it makes it easy for an agent to experience (and to think of) her prospected alternatives as ones that she can "get to" in the sense that libertarians want to pick out, and which they claim might be provided for by indeterminism. That is, it might easily seem that the future is open in a way that would require indeterminism for the experience to be accurate. Even so, all that is going on is that the agent is considering what outcomes she can cause, depending on which choices she makes, which requires letting the event of her choice range across more than one value in order to prospect the downstream effects of her choosing in different ways. Modeling her own choice in this way is no different in kind

from modeling how contingent events, such as different paths of an approaching hurricane, would cause various outcomes, such as damage to different cities.

Second, prospection models the event of choice as exogenous, and thus as carved off from its antecedent causes and allowed to vary freely across a range of values. If an agent experiences deliberation in this way, then she very likely experiences her choice as not having antecedent sufficient causes. And a choice that is experienced as not having antecedent sufficient causes is, a fortiori, experienced as not having antecedent deterministic causes.

Even so, belief in libertarian freedom is not justified by such experiences. For one thing, the hypotheticals generated in prospection are subjective and relative, and so they cannot support the view that the future is indeterministically open. This is partly due to the epistemic nature of the possibilities that such hypotheticals capture, and epistemic possibility is compatible with determinism. Moreover, even if one introspects that one's choice seems to lack deterministic causes, it does not follow from this that it actually lacks such causes, since many of the causes might not be introspectible. Finally, modeling choice in this way illuminates how the process of deliberation itself is a causal difference-maker—prospection, perhaps experienced as indeterministic, is an important causal contributor to what happens in the world, even if the world turns out to be deterministic.

In this way, my view provides a deflationary explanation of a major motivation for libertarianism—namely, the experience of deliberation and choice—while nevertheless granting that it is natural for agents to form the belief that their choice is indeterministic. At bottom, libertarian beliefs derive from agents' experiences of navigating into the future, rather than from their introspection on the causes of choices. Thus, people's indeterminist beliefs have their source in their practical experience of deliberating and choosing.

In Chapter 5, I also show how the sort of contextualism about the freedom to do otherwise that follows from my application of causal modeling to questions about modal freedom is distinct from the sort of contextualist proposal endorsed by Horgan (which I criticize in Chapter 3). I also show how my version of con-

textualism is immune to the specific criticism that I direct at Horgan's view. In short, consideration of Frankfurt-cases shows how worries about whether agents are free arise for Horgan's view even when contextual parameters are normal, whereas they do not arise for my view.

6.6 Conclusion

In the dissertation, I first present new empirical findings that support experience-incompatibilism on the descriptive question about modal freedom. Next, I defend these findings against an error theory for incompatibilist judgments about experiences of freedom developed by Terry Horgan. I then propose a compatibilist story about the veridicality of experiences of modal freedom that have libertarian content, and thus I defend experience-compatibilism on the prescriptive question—which asks how we *ought* to think about experiences of modal freedom. Finally, I develop a deflationary etiological story about the source of libertarian agentive experiences and beliefs.

What is at stake in all this is our self-image as agents who are free to navigate among alternative pathways into the future. The compatibilist view that I develop in the dissertation preserves this self-image, even assuming that determinism is true.

Bibliography

- [1] Ainge, J.A., M. Tamosiunaite, F. Worgotter, and P. A. Dudchencko. 2012. “Hippocampal Place Cells Encode Intended Destination, and Not a Discriminative Stimulus, in a Conditional T-maze Task,” *Hippocampus*, **22**: 534–543.
- [2] Anscombe, E. 1981/1971. “Causality and Determination,” in *Metaphysics and the Philosophy of Mind: Collected Philosophical Papers of G.E.M. Anscombe*. Minneapolis: University of Minnesota Press. 133–47.
- [3] Anscombe, G. E. M. 1957. *Intention*. Oxford: Blackwell.
- [4] Aristotle. ca. 330 BC/1999. *Nicomachean Ethics*, trans. T. Irwin. Indianapolis: Hackett.
- [5] Armstrong, D. 1983. *What Is a Law of Nature?* Cambridge: Cambridge University Press.
- [6] Armstrong, D. 1978. *A Theory of Universals*. Cambridge: Cambridge University Press.
- [7] Åström, K. J. and R. M. Murray. 2008. *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton: Princeton University Press.
- [8] Austin, J. L. 1956. “Ifs and Cans,” in *Philosophical Papers*, Third Edition. Oxford: Clarendon Press 1979: 205–32. Originally published in *Proceedings of the British Academy*, **42**.
- [9] Ayer, A. J. 1954. “Freedom and Necessity,” in *Philosophical Essays*. London and Basingstoke: Macmillan. 271–284.

- [10] Balaguer, M. 2010. *Free Will as an Open Scientific Problem*. Cambridge, MA: MIT Press.
- [11] Balaguer, M. 2004. "A Coherent, Naturalistic, and Plausible Formulation of Libertarian Free Will," *Noûs*, **38**(3): 379–406.
- [12] Barker, S. 2011. "Can Counterfactuals Really Be about Possible Worlds?" *Noûs*, **45**(3): 557–576.
- [13] Bayne, T. 2011. "The Sense of Agency," in F. Macpherson (ed.), *The Senses*. Oxford: Oxford University Press. 355–74.
- [14] Bayne, T. 2009. "Perception and the Reach of Phenomenal Content," *Philosophical Quarterly*, **59**(236): 385–404.
- [15] Bayne, T. 2008. "The Phenomenology of Agency," *Philosophy Compass*, **3**(1): 182–202.
- [16] Bayne, T. and E. Pacherie. 2007. "Narrators and Comparators: The Architecture of Agentive Self-Awareness," *Synthese*, **159**(3): 475–491.
- [17] Bealer, G. 1998. "Intuition and the Autonomy of Philosophy," in *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, M. R. DePaul and W. Ramsey (eds.), Lanham, Maryland: Rowman and Littlefield Publishers, Inc. 201–239.
- [18] Bedke, M. S. 2008. "Ethical Intuitions: What They Are, What They Are Not, and How They Justify," *American Philosophical Quarterly*, **45**(3): 253–270.
- [19] Bennett, J. 1984. "Counterfactuals and Temporal Direction," *The Philosophical Review*, **93**: 57–91.
- [20] Bernstein, M. 2002. "Fatalism," in R. Kane (ed.), *The Oxford Handbook of Free Will*. Oxford: Oxford University Press. 65–81.
- [21] Berofsky, B. 2002. "Ifs, Cans, and Free Will: The Issues," in R. Kane (ed.), *The Oxford Handbook of Free Will*. New York: Oxford University Press. 181–201.
- [22] Berridge, K. 2004. "Motivation Concepts in Behavioral Neuroscience," *Physiology and Behavior*, **81**: 179–209.

- [23] Bird, A. 2005. "The Dispositionalist Conception of Laws," *Foundations of Science*, **10**(4): 353–370.
- [24] Blakemore, S.-J. and C. Frith. 2003. "Self-Awareness and Action," *Current Opinion in Neurobiology*, Special Issue **13**(2): 219–224.
- [25] Brogaard, B. 2013. "The Phenomenal Use of 'Look'," *Philosophy Compass*.
- [26] Buckner, R. and D. Carroll. 2007. "Self-Projection and the Brain," *Trends in Cognitive Science*, **11**: 49–57.
- [27] Buhrmester, M.D., T. Kwang, and S.D. Gosling. 2011. "Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?" *Perspectives on Psychological Science*, **6**: 3-5.
- [28] Campbell, C. A. 1951. "Is 'Freewill' a Pseudo-Problem?" *Mind*, **60**(240): 441–65.
- [29] Campbell, John. 1997. "The Simple View of Colour," in A. Byrne and D. Hilbert (eds.), *Readings on Color*. MIT Press. 257–268.
- [30] Campbell, Joseph. 2005. "Compatibilist Alternatives," *Canadian Journal of Philosophy*, **35**:387–406.
- [31] Campbell, Joseph., M. O'Rourke, and D. Shier. 2004. "Freedom and Determinism: A Framework," in J. Campbell, M. O'Rourke, and D. Shier (eds.), *Freedom and Determinism*. Cambridge, MA: MIT Press, Bradford Books. 1–18.
- [32] Carroll, J. W. 2011. "Laws of Nature," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*.
- [33] Carver, C. S. and M. F. Scheier. 1990. "Origins and Functions of Positive and Negative Affect: A Control-Process View," *Psychological Review*, **97**: 19–35.
- [34] Chalmers, D. 2006. "Perception and the Fall from Eden," in T. Gendler and J. Hawthorne (eds.), *Perceptual Experience*. Oxford: Oxford University Press. 49–125.

- [35] Chalmers, D. 2004. "The Representational Character of Experience," in B. Leiter (ed.), *The Future for Philosophy*. Oxford: Oxford University Press. 153–181.
- [36] Chalmers, D. 2002. "On Sense and Intension," *Philosophical Perspectives*, 16. Atascadero: Ridgeview. 135–182.
- [37] Chisholm, R. 1957. *Perceiving: A Philosophical Study*. Ithaca: Cornell University Press.
- [38] Clarke, R. 2009. "Dispositions, Abilities to Act, and Free Will: The New Dispositionalism," *Mind*, 118: 323–51.
- [39] Conant, R. C. and W. R. Ashby. 1970. "Every Good Regulator of a System Must Be a Model of That System," *International Journal of Systems Science*, 1: 89–97.
- [40] Craig, A. D. 2009. "How Do You Feel—Now? The Anterior Insula and Human Awareness," *Nature Reviews Neuroscience*, 10: 59–70.
- [41] Danto, A. C. 1959. "The Paradigm Case Argument and the Free-Will Problem," *Ethics*, 69(2): 120–124.
- [42] Deery, O. Forthcoming. "Absences and Late Preemption," *Theoria*.
- [43] Deery, O., M. Bedke, and S. Nichols. 2013. "Phenomenal Abilities: Incompatibilism and the Experience of Agency," in David Shoemaker (ed.), *Oxford Studies in Agency and Responsibility*. New York: Oxford University Press. 126–150.
- [44] Deery, O., T. Davis, and J. Carey. In preparation. "The Free-Will Intuitions Scale and the Question of Natural Compatibilism."
- [45] Delk, J. L. and S. Fillenbaum. 1965. "Differences in Perceived Color as a Function of Characteristic Color," *The American Journal of Psychology*, 78(2): 290–93.
- [46] Dennett, D. 1978. "On Giving Libertarians What They Say They Want," in *Brainstorms*. Cambridge, MA: MIT Press.
- [47] Derdikman, D. and E. I. Moser. 2010. "A Manifold of Spatial Maps in the Brain," *Trends in Cognitive Sciences*, 14: 561–569.

- [48] Double, R. 1997. "Misdirection in the Free Will Problem," *American Philosophical Quarterly*, **34**(3): 357–366.
- [49] Double, R. 1996. *Metaphilosophy and Free Will*. New York: Oxford University Press.
- [50] Dretske, F. 1977. "Laws of Nature," *Philosophy of Science*, **44**: 248–268.
- [51] Dugatkin, L. A. 2004. *Principles of Animal Behavior*. New York: W. W. Norton.
- [52] Earman, J. 2004. "Determinism: What We Have Learned and What We Still Don't Know," in J. Campbell, M. O'Rourke, and D. Shier (eds.), *Freedom and Determinism*. Cambridge, MA: MIT Press, Bradford Books. 21–46.
- [53] Earman, J. 1992. *Bayes or Bust?* Cambridge, MA: MIT Press.
- [54] Earman, J. 1986. *A Primer on Determinism*. Dordrecht: D. Reidel Publishing Company.
- [55] Earman, J. 1984. "Laws of Nature: The Empiricist Challenge," in R. Bogdan (ed.), *D. M. Armstrong*. Dordrecht: D. Reidel Publishing Company.
- [56] Eykhoff, P. 1994. "Every Good Regulator of a System Must Be a Model of That System," *Modeling, Identification, and Control*, **15**: 135–139.
- [57] Fara, M. 2008. "Masked Abilities and Compatibilism," *Mind*, **117**(468): 843–865.
- [58] Fine, K. 2012. "A Difficulty for the Possible Worlds Analysis of Counterfactuals," *Synthese*, **189**(1): 29–57.
- [59] Fischer, J. M. 2013. "My Compatibilism," in P. Russell and O. Deery (eds.). 296–317.
- [60] Fischer, J. M. 2008. "Freedom, Foreknowledge, and Frankfurt: A Reply to Vihvelin," *Canadian Journal of Philosophy*, **38**(3): 327–42.
- [61] Fischer, J. M. 2007. "Compatibilism," in J. M. Fischer, R. Kane, D. Pereboom and M. Vargas, *Four Views on Free Will*. Oxford: Blackwell Publishing. 44–84.

- [62] Fischer, J. M. 1994. *The Metaphysics of Free Will*. Oxford: Blackwell Publishing.
- [63] Fischer, J. M. and M. Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- [64] Flew, A. 1959. "Divine Omniscience and Human Freedom," in A. Flew, and A. McIntyre (eds.), *New Essays in Philosophical Theology*. London: New York: Macmillan.
- [65] Frankfurt, H. 1969. "Alternate Possibilities and Moral Responsibility," *The Journal of Philosophy*, **66**(23): 829–39.
- [66] Gilbert, D. 2006. *Stumbling on Happiness*. New York: Knopf.
- [67] Gilbert, D. and T. Wilson. 2007. "Miswanting: Some Problems in the Forecasting of Future Affective States," in S. Lichtenstein and P. Slovic (eds.), *The Construction of Preferences*. Cambridge: Cambridge University Press. 550–563.
- [68] Gill, M. 1980. "Aristotle's Theory of Causal Action in Physics III, 31," *Phronesis* **25**(2): 129–147.
- [69] Ginet, C. 1997. "Freedom, Responsibility, and Agency," in *The Journal of Ethics*, **1**(1): 85–98.
- [70] Grabenhorst, F. and E. T. Rolls. 2011. "Value, Pleasure, and Choice in the Ventral Prefrontal Cortex," *Trends in Cognitive Sciences*, **15**: 56–67.
- [71] Grünbaum, A. 1952. "Causality and the Science of Human Behavior," *American Scientist*, **40**(4): 665–676.
- [72] Gupta, A. S., M. A. A. van der Meer, D. S. Touretzky, and A. D. Redish. 2010. "Hippocampal Replay is Not a Simple Function of Experience," *Neuron*, **65**: 695–705.
- [73] Hanson, R. N. 1958. *Patterns of Discovery*. Cambridge: Cambridge University Press.
- [74] Heller, M. 1996. "The Mad Scientist Meets the Robot Cats: Compatibilism, Kinds, and Counterexamples," *Philosophy and Phenomenological Research*, **56**(2): 333–37.

- [75] Hitchcock, C. R. 2001. "The Intransitivity of Causation Revealed in Equations and Graphs," *Journal of Philosophy*, **98**: 273–299.
- [76] Hobart, R. E. 1934. "Free Will as Involving Determination and Inconceivable without It," *Mind*, **43**(169): 1–27.
- [77] Hobbs, J. Unpublished. *What Can Ought Imply?* Ph.D. thesis, Cornell University.
- [78] Hofer, C. 2010. "Causal Determinism," in E. N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*.
- [79] Holton, R. 2010. "Disentangling the Will," in R. Baumeister, A. Mele and K. Vohs (eds.), *Free will and Consciousness: How Might They Work?* New York: Oxford University Press. 82–100.
- [80] Holton, R. 2006. "The Act of Choice," *Philosophers' Imprint*, **6**: 1–15.
- [81] Honderich, T. 1993. *How Free are You? The Determinism Problem*. Oxford: Oxford University Press.
- [82] Honoré, A. M. 1964. "Can and Can't," *Mind*, **73**: 463–479.
- [83] Horgan, T. Forthcoming. "Causal Compatibilism about Agentive Phenomenology," in M. Sabates, D. Sosa, and T. Horgan (eds.), *Festschrift for J. Kim*.
- [84] Horgan, T. 2012. "Introspection about Phenomenal Consciousness: Running the Gamut from Infallibility to Impotence," in D. Smithies and D. Stoljar (eds.), *Introspection and Consciousness*. Oxford: Oxford University Press. 403–422.
- [85] Horgan, T. 2011. "The Phenomenology of Agency and Freedom: Lessons from Introspection and Lessons from Its Limits," *Humana.Mente*, **15**: 77–97.
- [86] Horgan, T. 2007. "Agentive Phenomenal Intentionality and the Limits of Introspection," *Psyche*, **13**(2): 1–29.
- [87] Horgan, T and U. Kriegel. 2008. "Phenomenal Intentionality Meets the Extended Mind," *The Monist*, **91**: 353–380.

- [88] Horgan, T. and J. Tienson, 2002. "The Intentionality of Phenomenology and the Phenomenology of Intentionality," in D. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford: Oxford University Press. 520–33.
- [89] Hume, D. 1960/1739. *A Treatise on Human Nature*, L. A. Selby-Bigge (ed.). Oxford: Clarendon Press.
- [90] Ismael, J. 2013. "Causation, Free Will, and Naturalism," in H. Kincaid, J. Ladyman, and D. Ross (eds.), *Scientific Metaphysics*. New York: Oxford University Press. 208–235.
- [91] Johnston, M. 1992. "How to Speak of the Colors," *Philosophical Studies*, **68**(3): 221–263.
- [92] Jenkins, C. S. I. Forthcoming. "'Intuition,' Intuition, Concepts and the A Priori," in A. Booth and D. Rowbottom (eds.), *Intuitions*. Oxford University Press.
- [93] Kapitan, T. 1986. "Deliberation and the Presumption of Open Alternatives," *The Philosophical Quarterly*, **36**(143): 230–251.
- [94] Kratzer, A. 1977. "What 'Must' and 'Can' Must and Can Mean," *Linguistics and Philosophy*, **1**(3): 337–355.
- [95] Kriegel, U. Forthcoming. "A Hesitant Defense of Introspection," *Philosophical Studies*.
- [96] Kriegel, U., 2002. "Phenomenal Content," *Erkenntnis*, **57**: 175–198.
- [97] Kringelbach, M. and K. Berridge. 2009. "Towards a Functional Neuroanatomy of Pleasure and Happiness," *Trends in Cognitive Science*, **13**: 479–487.
- [98] Lehrer, K. 1960. "Can We Know that We Have Free Will by Introspection?" *The Journal of Philosophy*, **57**(5): 145–57.
- [99] Lehrer, K. 1968. "Cans Without Ifs," *Analysis*, **29**(1): 29–32.
- [100] Lehrer, K. 1976. "'Can' in Theory and Practice: A Possible World Analysis," in Brand and Walton (eds.), *Action Theory*. Dordrecht: D. Reidel. 241–270.

- [101] Levin, D. T. and M. R. Banaji. 2006. "Distortions in the Perceived Lightness of Faces: The Role of Race Categories," *Journal of Experimental Psychology: General*, **135**(4): 501–12.
- [102] Lewis, D. 1994. "Humean Supervenience Debugged," *Mind*, **103**: 473–490
- [103] Lewis, D. 1986. "The Paradoxes of Time Travel," in *Philosophical Papers, Vol. II*. Oxford: Oxford University Press.
- [104] Lewis, D. 1983. "New Work for a Theory of Universals," *Australasian Journal of Philosophy*, **61**(4): 343–377.
- [105] Lewis, D. 1981. "Are We Free to Break the Laws?" *Theoria*, **47**: 112–121.
- [106] Lewis, D. 1973. *Counterfactuals*. Harvard: Harvard University Press.
- [107] Lycan, W. 1996. *Consciousness and Experience*. Cambridge, MA: MIT Press.
- [108] MacPherson, F. 2012. "Cognitive Penetration of Colour Experience: Rethinking the Issue in Light of an Indirect Mechanism," *Philosophy and Phenomenological Research*, **84**(1): 24–62.
- [109] Mele, A. 2006. *Free Will and Luck*. New York: Oxford University Press.
- [110] Mele, A. 2003. "Agents' Abilities," *Noûs*, **37**: 447–470.
- [111] Mele, A. 1995. *Autonomous Agents*. New York: Oxford University Press.
- [112] Mill, J. S. 1865. *An Examination of William Hamilton's Philosophy*. Boston: William V. Spencer.
- [113] Miller, G., E. Galanter, and K. Pribram. 1960. *Plans and the Structure of Behavior*. N.Y.: Holt, Rhinehart, and Winston.
- [114] Monroe, A. E. and B. F. Malle. 2010. "From Uncaused Will to Conscious Choice: The Need to Study, Not Speculate about People's Folk Concept of Free Will," *Review of Philosophy and Psychology*, **9**: 211–224.
- [115] Moore, G. E. 1912. "Free Will," from *Ethics*. Oxford: Oxford University Press. 122–37.

- [116] Nahmias, E., S. Morris, T. Nadelhoffer and J. Turner. 2004. "The Phenomenology of Free Will," *Journal of Consciousness Studies*, **11**(7–8): 162–79.
- [117] Nahmias, E. and D. Murray. 2011. "Experimental Philosophy on Free Will: An Error Theory for Incompatibilist Intuitions," in *New Waves in Philosophy of Action*, Aguilar, J., A. Buckareff, and K. Frankish (eds.). Palgrave Macmillan. 189-216.
- [118] Nelkin, D. 2004. "Deliberative Alternatives," *Philosophical Topics*, **32**: 215–240.
- [119] Nichols, S. 2012. "The Indeterminist Intuition: Sources and Status," *The Monist*, **95**(2): 290–307.
- [120] Nichols, S. 2008. "How Can Psychology Contribute to the Free Will Debate?" in J. Baer, J. C. Kaufman and R. F. Baumeister (eds.), *Are We Free? Psychology and Free Will*. New York: Oxford University Press. 10–31.
- [121] Nichols, S. 2004. "The Folk Psychology of Free Will: Fits and Starts," *Mind and Language*, **19**(5): 473–502.
- [122] Nichols, S. and Joshua Knobe. 2007. "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions," *Noûs*, **41**(4): 663–685.
- [123] O'Connor, T. 1995. "Agent Causation," in T. O'Connor (ed.), *Agents, Causes, and Events: Essays on Indeterminism and Free Will*. New York: Oxford University Press. 173–200.
- [124] Palmer, S. E. 1999. *Vision Science: Photons to Phenomenology*. Cambridge, MA: MIT Press.
- [125] Paul, L. A. 2010. "A New Role for Experimental Work in Metaphysics," *Review of Philosophical Psychology*, **1**(3): 461–476.
- [126] Parfit, D. 1987. "Divided Minds and the Nature of Persons," in C. Blakemore and S. Greenfield (eds.), *Mindwaves: Thoughts on Intelligence, Identity and Consciousness*. Oxford: Blackwell Press.

- [127] Parfit, D. 1971. "Personal Identity," *The Philosophical Review*, **80**(1): 3–27.
- [128] Pearl, J. 2000. *Causality*. Oxford: Oxford University Press.
- [129] Pereboom, D. 2011. *Consciousness and the Prospects of Physicalism*. New York: Oxford University Press.
- [130] Pereboom, D. 2008. "A Compatibilist Account of the Epistemic Conditions on Rational Deliberation," *The Journal of Ethics*, **12**(3): 287–306.
- [131] Perry, J. 2004. "Compatibilist Options," in J. Campbell, M. O'Rourke, and D. Shier (eds.), *Freedom and Determinism*. MIT Press. 231–254.
- [132] Preuschoff, K., P. Bossaerts, and S. Quartz. 2006. "Neural Differentiation of Expected Reward and Risk in Human Subcortical Structures," *Neuron*, **51**: 381–390.
- [133] Quartz, S. R. 2009. "Reason, Emotion, and Decision-Making: Risk and Reward Computation with Feeling," *Trends in Cognitive Sciences*, **13**: 209–215.
- [134] Railton, P. In Press. "The Affective Dog and Its Rational Tale," *Ethics*.
- [135] Railton, P. 2002. "Kant Meets Aristotle Where Reason Meets Appetite," in C. U. Moulines and K.-G. Niebergall (eds.), *Argument Und Analyse*. Paderborn, Germany: Mentis.
- [136] Reid, T. 1788. *Essays on the Powers of the Human Mind*. London: Tegg, Griffin & Cumming.
- [137] Rolls, E. T., J. Tromans and S. M. Stringer. 2008. "Spatial Scene Representations Formed by Self-Organizing Learning in a Hippocampal Extension of the Ventral Visual System," *European Journal of Neuroscience*, **28**: 2116–2127.
- [138] Rosenthal, D. 1990. "A Theory of Consciousness," in N. Block, O. Flanagan, and G. Güzeldere (eds.), *The Nature of Consciousness*. Cambridge, MA: MIT Press. 773–788.

- [139] Russell, P. and O. Deery (eds.). 2013. *The Philosophy of Free Will: Essential Readings from the Contemporary Debates*. New York: Oxford University Press.
- [140] Saunders, J. T. 1968. "The Temptations of Powerlessness," *American Philosophical Quarterly*, **5**: 100–108.
- [141] Searle, J. 1984. *Minds, Brains, and Science*. Cambridge, MA: Harvard UP.
- [142] Seligman, M., P. Railton, R. Baumeister, and C. Sripada. 2013. "Navigating into the Future or Driven by the Past: Prospection as an Organizing Principle of Mind," *Perspectives on Psychological Science*, **8**(2): 119–141.
- [143] Shariff, A. F., J. D. Greene, and J. W. Schooler. Forthcoming. "His Brain Made Him Do It: Encouraging a Mechanistic Worldview Reduces Punishment."
- [144] Siegel, S. 2006. "Which Properties are Represented in Perception?" in T. Gendler Szabo and J. Hawthorne (eds.), *Perceptual Experience*. Oxford: Oxford University Press. 481–503.
- [145] Singer, T., H. D. Critchley, and K. Preuschoff. 2009. "A Common Role of Insula in Feelings, Empathy and Uncertainty," *Trends in Cognitive Sciences*, **13**: 334–340.
- [146] Smart, J. J. C. 1961. "Free Will, Praise and Blame," *Mind*, **70**(2): 91–306.
- [147] Smith, M. 2004. "Rational Capacities," in *Ethics and the A Priori: Selected Essays on Moral Psychology and Meta-Ethics* by M. Smith. New York: Cambridge University Press. 114–135.
- [148] Schultz, W. 2002. "Multiple Reward Signals in the Brain," *Nature Reviews Neuroscience*, **1**: 199–207.
- [149] Strawson, G. 1994. "The Impossibility of Moral Responsibility," *Philosophical Studies*, **75**(1–2): 5–24.
- [150] Symons, J. 2008. "Intuition and Philosophical Methodology," *Axiomathes*, **18**(1): 67–89.

- [151] Synofzik, M., G. Vosgerau and A. Newen. 2008. "Beyond the Comparator Model: A Multi-Factorial Two-Step Account of Agency," *Consciousness and Cognition*, **17**(1): 219–239.
- [152] Taylor, R. 1983. *Metaphysics, 3rd edition*. Englewood Cliffs, N.J.: Prentice Hall.
- [153] Tobler, D. N., J. P. Dougherty, R. J. Dolan, W. Schultz. 2006. "Reward Value Coding Distinct from Risk Attitude-Related Uncertainty Coding in Human Reward Systems," *Journal of Neurophysiology*, **97**: 1621–1632.
- [154] Tooley, M. 1977. "The Nature of Laws," *Canadian Journal of Philosophy*, **7**: 667–698.
- [155] Turner, J. and E. Nahmias. 2006. "Are the Folk Agent-Causationists?" *Mind and Language*, **21**(5): 597–609.
- [156] Tye, M. 1995. *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: MIT Press.
- [157] van Inwagen, P. 2008. "How To Think About the Problem of Free Will," *The Journal of Ethics*, **12**(3): 327–341.
- [158] van Inwagen, P. 2000. "Free Will Remains a Mystery," *Philosophical Perspectives*, **14**: 1–20.
- [159] van Inwagen, P. 1983. *An Essay on Free Will*. Oxford: Oxford University Press.
- [160] Vargas, M. 2013. "How to Solve the Problem of Free Will," in P. Russell and O. Deery (eds.), 400–416.
- [161] Velleman, J. D. 2006. "So It Goes," The Amherst Lecture in Philosophy.
- [162] Vihvelin, K. 2011. "Arguments for Incompatibilism," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*.
- [163] Vihvelin, K. 2004. "Free Will Demystified: A Dispositional Account," *Philosophical Topics*, **32**(1–2): 427–450.
- [164] Vohs, K. D., and J. W. Schooler. 2008. "The Value of Believing in Free Will: Encouraging a Belief in Determinism Increases Cheating," *Psychological Science*, **19**: 49–54.

- [165] Widerker, D. 2006. "Blameworthiness and Frankfurt's Argument Against the Principle of Alternative Possibilities," in D. Widerker and M. McKenna (eds.), *Moral Responsibility and Alternative Possibilities*. Aldershot: Ashgate Publishing. 53–73.
- [166] Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press.

Appendix A

Survey Materials for Chapter 2

STUDY 1

CONDITION 1

[Page 1]

Thank you for consenting to participate in the study.

On the next page, you will read a brief passage. Then we will ask you some questions.

Click “Next” below to begin.

[Go to Page 2]

[Page 2]

Please read the following passage, and answer the questions that follow as best you can:

Imagine that you are sledding down a snowy path on a mountainside. Your sled has a steering mechanism that allows you to control the direction of the sled.

Below you is a fork in the path with snow built up in the middle, and you can tell that, if you don't direct your sled one way or the other, the contours of the mountain will channel you and your sled either to the left or to the right.

Consider how things seem to you as you approach the fork in the path. In particular, consider what it's like to decide which way the sled will go.

Please indicate your level of agreement with the following statement:

When deciding which way the sled will go, it feels like I can either go to the left or go to the right.

1. Disagree completely
2. Strongly disagree
3. Disagree
4. Neither agree nor disagree
5. Agree
6. Strongly agree
7. Agree completely¹

[If 1–4, then go to Page 3; if 5–7, then go to Page 5]

[Page 3]

Still considering how things seem to you as you approach the fork in the path, please indicate your level of agreement with the following statement:

When deciding which way the sled will go, it feels like I cannot either go to the left or go to the right.

[If 1–4, then go to Page 4; if 5–7, then go to Page 12]

¹ This 7-point Likert scale was used whenever participants were asked to indicate their level of agreement with a statement. It will be omitted from the materials presented below, and should be taken for granted.

[Page 4]

Still considering how things seem to you as you approach the fork in the path, please indicate your level of agreement with the following statement:

When deciding which way the sled will go, it is unclear whether or not the following is true: it feels like I can either go to the left or go to the right.

[Go to Page 12]

[Page 5]

We will now describe to you the notion of causal completeness and then ask you how it relates to the sledding situation.

According to causal completeness, everything that happens is fully caused by what happened before it. This is true from the very beginning of the universe, so what happened in the beginning of the universe fully caused what happened next, and so on right up until the present. Causal completeness holds that everything is fully caused in this way, including people's decisions.

To understand this idea, let's start with an example of a real event. On May 18, 1980, Mount St. Helens erupted into a fiery volcano. The eruption was triggered by a 5.1 earthquake. Scientists agree that this earthquake (along with other factors) fully caused the eruption.

According to causal completeness, if we could somehow replay the entire past right up until St. Helens erupted on May 18, 1980, then St. Helens would once again erupt at that time. Another way to put this is to say that all the events leading up to the eruption made it so that the eruption had to happen.

The causal completeness view maintains that this is true for everything in the universe, not just volcanoes and earthquakes. So, to take another real example, on August 22, 2008, Barack Obama decided to have Joe Biden as his Vice Presidential running mate in the U.S. Presidential election. There were obviously many factors that led up to this decision—including Obamas feelings about Biden, his

feelings about Hillary Clinton, his beliefs about how each would help his candidacy, his beliefs about how each would actually function as Vice President, and so on. According to causal completeness, if we replayed the past right up until Obama's decision—including everything that was going through Obama's mind—then Obama would once again make exactly the same decision. That is, all the events leading up to Obama's decision (including everything that was going through Obama's mind), made it so that it had to happen that Obama would pick Biden.

Causal completeness also maintains that this goes for everything in the future. So in the future, every time you make a decision, what you ultimately decide will be fully caused by everything that happens leading up to the decision. That is, given everything that happens before any future decision of yours, that future decision has to happen the way that it does.

Just to be sure that you understand the idea of causal completeness, please indicate whether the following is True or False.

According to causal completeness, St. Helens would have erupted on May 18, 1980 even if there had been no earthquake.

True / False

[If true, then go to Page 6; if false, then go to Page 7]

[Page 6]

The correct answer is False.

Recall that causal completeness says that what happens is fully caused by what happened before it. So causal completeness doesn't say that the eruption of St. Helens was inevitable no matter what. Rather, causal completeness would say that the eruption happened because of the earthquake.

Given this clarification, please indicate whether the following statement is True or False.

According to causal completeness, if there hadn't been an earthquake, St. Helens still would have erupted on May 18, 1980.

True / False

[If true, then go to Page 12; if false, then go to Page 7]

[Page 7]

Just one more question to make sure you understand causal completeness. Please indicate whether the following statement is True or False.

According to causal completeness, if a week from now Barack Obama decides to have soda with dinner, all the events leading up to that decision will make it the case that he has to decide to have a soda with dinner.

True / False

[If true, then go to Page 8; if false, then go to Page 12]

[Page 8]

The correct answer is True.

Recall that causal completeness says that what happens in the future is fully caused by what happened in the past. There is no exception for human decisions. According to causal completeness, when you make a decision in the future, that decision will be fully caused by everything leading up to it. So causal completeness says that if you make a decision in the future—for example, if a week from now you decide to have soda with dinner—that decision has to happen in the particular way that it does, given everything that leads up to it (including everything that is going through your mind right before the decision).

Given this clarification, please indicate whether the following statement is True or False.

According to causal completeness, if everything that happens leading up to Obama's future decision—for example, his decision a week from now to have

soda with dinner—stays exactly the same, then that decision does not have to happen.

True / False

[If true, then go to Page 12; if false, then go to Page 9]

[Page 9]

Now, recall the sledding situation. You previously agreed with the following statement:

When deciding which way the sled will go, it feels like I can either go to the left or go to the right.

Considering this previous statement and your understanding of causal completeness, please indicate your level of agreement with the following:

Even though it felt like I could either go to the left or go to the right, if causal completeness is true there is something mistaken about how that decision felt to me.

[Go to Page 10]

[Page 10]

Please briefly explain why you answered the last question as you did. (This question is optional.)

[Go to Page 11]

[Page 11]

Thank you for completing the main portion of the survey. If you have any comments about the questions, please enter them below. Otherwise, please help us out by entering some demographic information in the few questions that follow before finalizing your survey.

[Go to page Page 12]

[Page 12]

[The following demographic information was gathered: participants' sex, age, the number of philosophy classes they had taken, whether participants had taken a class in which free will was a topic, participants' ethnic background, religious affiliation, and level of religiosity.]

STUDY 1

CONDITION 2

[This condition used the same introduction page and training section, Pages 5–8, as Study 1, Condition 1.]

Page 2

Please read the following passage, and answer the questions that follow as best you can:

Imagine that, many years ago, you were sledding down a snowy path on a mountainside. Your sled had a steering mechanism that allowed you to control the direction of the sled. Below you was a fork in the path with snow built up in the middle, and you could tell that, if you didn't direct your sled one way or the other, the contours of the mountain would channel you and your sled either to the left or to the right. In the end, you decided to go left and you went left.

Still imagining that you made this decision many years ago, and assuming no extra facts not described in the passage above, please indicate your level of agreement with the following statement:

I could have gone right instead of left.

[If 1–4, then go to Page 3; if 5–7, then go to Page 5]

[Page 3]

Still imagining that you made this decision many years ago, and assuming no extra facts not described in the passage above, please indicate your level of agreement with the following statement:

I could not have gone right instead of left.

[If 1–4, then go to Page 4; if 5–7, then go to Page 12]

[Page 4]

Still imagining that you made this decision many years ago, and assuming no extra facts not described in the passage above, please indicate your level of agreement with the following statement:

It is unclear whether or not the following is true: I could have gone right instead of left.

[Go to Page 12]

[Page 9]

Now, recall the sledding situation. You previously agreed with the following statement:

I could have gone right instead of left.

Considering this previous statement and your understanding of causal completeness, please indicate your level of agreement with the following:

Even though I said I could have gone right instead of left, if causal completeness is true there is something mistaken about what I said.

[Go to Page 10]

[Page 10]

Please briefly explain why you answered the last question as you did. (This question is optional.)

[Go to Page 11]

[Page 11]

Thank you for completing the main portion of the survey. If you have any comments about the questions, please enter them below. Otherwise, please help us

out by entering some demographic information in the few questions that follow before finalizing your survey.

[Go to page Page 12]

STUDY 2

[This study used the same introduction page and training section, Pages 5–8, as Study 1, Condition 1.]

CONDITION 1

[Page 2]

Please read the following passage, and answer the questions that follow as best you can:

Imagine that you have \$0.50 to donate. You have two options:

Donate to a foundation that protects the endangered tree *castanea dentata*.

OR

Donate to a foundation that protects the endangered tree *ulmus dentata*.

These are your only two options.

Now, consider how things seem to you as you face your decision. In particular, consider what it's like to decide which option to choose.

Please indicate your level of agreement with the following statement:

When deciding which option to choose, it feels like I can either choose to donate to *castanea dentata* or choose to donate to *ulmus dentata*.

[If 1–4, then go to Page 3; if 5–7, then go to Page 5]

[Page 3]

Still considering how things seem to you as you face your decision, please indicate your level of agreement with the following statement:

When deciding which option to choose, it feels like I cannot either choose to donate to *castanea dentata* or choose to donate to *ulmus dentata*.

[If 1–4, then go to Page 4; if 5–7, then go to Page 12]

[Page 4]

Still considering how things seem to you as you face your decision, please indicate your level of agreement with the following statement:

When deciding which option to choose, it is unclear whether or not the following is true: it feels like I can either choose to donate to *castanea dentata* or choose to donate to *ulmus dentata*.

[Go to Page 12]

[Page 9]

Now, recall the donation situation. You previously agreed with the following statement:

When deciding which option to choose, it feels like I can either choose to donate to *castanea dentata* or choose to donate to *ulmus dentata*.

Considering this previous statement and your understanding of causal completeness, please indicate your level of agreement with the following:

Even though it felt like I could either choose to donate to *castanea dentata* or choose to donate to *ulmus dentata*, if causal completeness is true then I couldn't really have chosen differently than I did.

[Go to Page 10]

STUDY 2

CONDITION 2

[Page 2]

Please read the following passage, and answer the questions that follow as best you can:

You have \$0.50 to donate. We, the researchers, will actually donate this money for you whichever way you decide. You have two options:

Donate to a foundation that protects the endangered tree *castanea dentata*.

OR

Donate to a foundation that protects the endangered tree *ulmus dentata*.

These are your only two options. Each option is currently available for you to choose at the bottom of this page. But don't decide just yet.

First, consider how things seem to you as you face your decision. In particular, consider what it's like to decide which option to choose.

Please indicate your level of agreement with the following statement:

When deciding which option to choose, it feels like I can either choose to donate to *castanea dentata* or choose to donate to *ulmus dentata*.

HERE ARE YOUR TWO OPTIONS:

After you have answered the question above, please choose a charity:

The endangered tree *castanea dentata* / The endangered tree *ulmus dentata*

[If 1–4, then go to Page 3; if 5–7, then go to Page 5]

[Page 3]

Considering how things seemed to you as you faced your decision, please indicate your level of agreement with the following statement:

When deciding which option to choose, it felt like I couldn't either choose to donate to *castanea dentata* or choose to donate to *ulmus dentata*.

[If 1–4, then go to Page 4; if 5–7, then go to Page 12]

[Page 4]

Still considering how things seemed to you as you faced your decision, please indicate your level of agreement with the following statement:

When deciding which option to choose, it was unclear whether or not the following was true: it felt like I could either choose to donate to *castanea dentata* or choose to donate to *ulmus dentata*.

[Go to Page 12]

[Page 9]

Now, recall the donation situation. You previously agreed with the following statement:

When deciding which option to choose, it feels like I can either choose to donate to *castanea dentata* or choose to donate to *ulmus dentata*.

Considering this previous statement and your understanding of causal completeness, please indicate your level of agreement with the following:

Even though it felt like I could either choose to donate to *castanea dentata* or choose to donate to *ulmus dentata*, if causal completeness is true then I couldn't really have chosen differently than I did.

[Go to Page 10]

STUDY 2

CONDITION 3

[Page 2]

Please read the following passage, and answer the questions that follow as best you can:

You have \$0.50 to donate. We, the researchers, will actually donate this money for you whichever way you decide. You have two options:

Donate to a foundation that protects the endangered tree *castanea dentata*.

OR

Donate to the Childhood Cancer Foundation.

These are your only two options. Each option is currently available for you to choose at the bottom of this page. But don't decide just yet.

First, consider how things seem to you as you face your decision. In particular, consider what it's like to decide which option to choose.

Please indicate your level of agreement with the following statement:

When deciding which option to choose, it feels like I can either choose to donate to the endangered tree *castanea dentata* or choose to donate to the Childhood Cancer Foundation.

HERE ARE YOUR TWO OPTIONS:

After you have answered the question above, please choose a charity:

The endangered tree *castanea dentata* / The Childhood Cancer Foundation

[If 1–4, then go to Page 3; if 5–7, then go to Page 5]

[Page 3]

Considering how things seemed to you as you faced your decision, please indicate your level of agreement with the following statement:

When deciding which option to choose, it felt like I couldn't either choose to donate to the endangered tree *castanea dentata* or choose to donate to the Childhood Cancer Foundation.

[If 1–4, then go to Page 4; if 5–7, then go to Page 12]

[Page 4]

Still considering how things seemed to you as you faced your decision, please indicate your level of agreement with the following statement:

When deciding which option to choose, it was unclear whether or not the following was true: it felt like I could either choose to donate to the endangered tree *castanea dentata* or choose to donate to the Childhood Cancer Foundation.

[Go to Page 12]

[Page 9]

Now, recall the donation situation. You previously agreed with the following statement:

When deciding which option to choose, it feels like I can either choose to donate to the endangered tree *castanea dentata* or choose to donate to the Childhood Cancer Foundation.

Considering this previous statement and your understanding of causal completeness, please indicate your level of agreement with the following:

Even though it felt like I could either choose to donate to the endangered tree *castanea dentata* or choose to donate to the Childhood Cancer Foundation, if

causal completeness is true then I couldn't really have chosen differently than I did.

[Go to Page 10]

STUDY 3

[This study used the same introduction page and training section, Pages 5–8, as Study 1, Condition 1.]

CONDITION 1

[Page 2]

Please read the following passage, and answer the questions that follow as best you can:

At the bottom of this page, there are two buttons, labelled H and V. Each option is currently available for you to choose. In a moment, we'll ask you to choose just one of them. For this survey, only one of the buttons will give you an extra \$0.05 (as bonus payment on MTurk) if you choose it. But we won't tell you which button it is—you'll have to make a choice and find out.

But don't decide just yet.

First, consider how things seem to you as you face your decision. In particular, consider what it's like to decide which option to choose.

Please indicate your level of agreement with the following statement:

When deciding which option to choose, it feels like I can either choose H or choose V.

HERE ARE YOUR TWO OPTIONS:

After you have answered the question above, please choose a button.

Once you have you chosen, we'll ask you a few more questions before telling you whether you picked the bonus button.

H / V

[If 1–4, then go to Page 3; if 5–7, then go to Page 5]

[Page 3]

Considering how things seem to you as you face your decision, please indicate your level of agreement with the following statement:

When deciding which option to choose, it feels like I can either choose H or choose V.

[If 1–4, then go to Page 4; if 5–7, then go to Page 12]

[Page 4]

Still considering how things seem to you as you face your decision, please indicate your level of agreement with the following statement:?

When deciding which option to choose, it is unclear whether or not the following is true: it feels like I can either choose H or choose V.

[Go to Page 12]

[Page 9]

Now, recall the button-choosing situation. You previously agreed with the following statement:

When deciding which option to choose, it feels like I can either choose H or choose V.

Considering this previous statement about how things felt to you before your choice and your understanding of causal completeness, please indicate your level of agreement with the following:?

If causal completeness is true, then I couldn't really have chosen differently than I did.

[Go to Page 10]

STUDY 3

CONDITION 2

[Page 2]

Please read the following passage, and answer the questions that follow as best you can:

At the bottom of this page, there are two buttons, labelled H and V. Each option is currently available for you to choose. In a moment, we'll ask you to choose just one of them. For this survey, only one of the buttons will give you an extra \$0.05 (as bonus payment on MTurk) if you choose it. But we won't tell you which button it is—you'll have to make a choice and find out.

But don't decide just yet.

First, consider how things seem to you as you face your decision. In particular, consider what it's like to wonder which option you'll choose.

Please indicate your level of agreement with the following statement:

When wondering which option I'll choose, it feels like I don't know for sure before I select a button which button is the bonus button.

HERE ARE YOUR TWO OPTIONS:

After you have answered the question above, please choose a button.

Once you have you chosen, we'll ask you a few more questions before telling you whether you picked the bonus button.

H / V

[If 1–4, then go to Page 3; if 5–7, then go to Page 5]

[Page 3]

Considering how things seem to you as you face your decision, please indicate your level of agreement with the following statement:

When wondering which option I'll choose, it feels like I do know for sure before I select a button which button is the bonus button.

[If 1–4, then go to Page 4; if 5–7, then go to Page 12]

[Page 4]

Still considering how things seem to you as you face your decision, please indicate your level of agreement with the following statement:

When wondering which option I'll choose, it is unclear whether or not the following is true: it feels like I don't know for sure before I select a button which button is the bonus button.

[Go to Page 12]

[Page 9]

Now, recall the button-choosing situation. You previously agreed with the following statement:

When wondering which option I'll choose, it feels like I don't know for sure before I select a button which button is the bonus button.

Considering this previous statement about how things felt to you before your choice and your understanding of causal completeness, please indicate your level of agreement with the following:

If causal completeness is true, then I knew for sure before I selected a button which button was the bonus button.

[Go to Page 10]