IS COMMON SENSE NEITHER COMPATIBILIST NOR INCOMPATIBILIST?

BY

DAN SIMPSON BECKLLOYD

A thesis

submitted to the Victoria University of Wellington

in fulfilment of the requirements for the degree of

Master of Arts in Philosophy

Victoria University of Wellington

2021

Table of Contents

Abstract

In the debate between the compatibilist and the incompatibilist, there has been significant dispute about who has the burden of proof. While both sides often agree that the burden lies on those who argue against commonsense belief, they disagree on what those beliefs actually are. Kadri Vihvelin takes the rare position that there is actually no commonsense belief about the compatibility question. This is derived from the claim that there is no commonsense belief about whether or not determinism is true. And from this, Vihvelin concludes that both sides have an argumentative burden. She then applies a burden according to the modal claims made by each side of the debate, issuing the verdict that the higher burden is on the incompatibilist because they make an impossibility claim. Though Vihvelin clearly makes empirical claims about commonsense beliefs related to free will and determinism, she also presents a critique of philosophical intuitions that suggests scepticism with regard to empirical results bearing on such beliefs. She suggests that the materials may produce intuitions that do not reflect beliefs held prior to the experiment. But this betrays a dilemma: either we can use experiments to answer these empirical questions, in which case we should look at the best available evidence, or we can't use experiments to answer such questions, in which case we should remain silent on them. Ultimately, the current state of research does provide answers, albeit incomplete. There is still room for improved materials and wider studies, but there is nonetheless strong evidence against Vihvelin's empirical claims. This has implications not only for Vihvelin's arguments, but for burden of proof claims regarding the compatibilism/incompatibilism debate more generally and emphasises the need for philosophers of free will to take the empirical results more seriously.

Acknowledgements

Chapter 1: Introduction

When debating the compatibility of free will and determinism, philosophers have long disagreed about which side of the debate has the burden of proof. Many such arguments have appealed to whatever it is that folk commonly believe, whether that be what "ordinary persons" believe (Kane, 1999, p. 218), our "pretheoretic" beliefs (Ekstrom, 2005, p. 310), "our general, unreflective, implicit, non-philosophical conception" (Strawson, 2010, p. 26), what "we naturally tend to assume" (Pink, 2004, pp. 12-13), among other phrases with similar connotations. Kadri Vihvelin (2013), in *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*, appeals to "common sense". The back and forth is usually between those who claim the compatibilist has the burden and those who claim it is on the incompatibilist. Vihvelin stakes out the rare position of arguing that commonsense beliefs are *neither* compatibilist nor incompatibilist, so *neither* has *the* burden of proof – rather, *both* must argue their cases. She goes further, arguing that since the incompatibilist makes an impossibility claim where the compatibilist does not, the argumentative burden is higher for the incompatibilist (p. 30).

Supposing that the burden of proof should indeed lie with those who argue against the common view, and assuming there actually *is* a common view, it strikes me that the most definitive way of establishing where that burden lies is to try and capture this common view empirically. Research of this kind began in experimental psychology some 60 years ago though the bulk of the research was done in the last 20 years as part of the experimental philosophy movement. On the face of it then, contemporary burden of proof arguments should be quite straightforward: the evidence indicates that the common view is x, therefore the burden of proof is on those denying x.

However, an aspect of Vihvelin's argument suggests that this might not be so straightforward. She presents a scepticism of the kinds of intuitions produced by thought experiments that calls into question whether empirical research can get at folk pre-theoretic beliefs at all. But this betrays a dilemma. Either it is possible to capture commonsense beliefs about free will and determinism through empirical means or it is not. If it is, then we should use such means to resolve the question of who has the burden of proof. If it is not possible, then we have no basis from which to make claims about these commonsense beliefs.

Speculation is an option, but your speculation is just as good as mine, so the debate seems moot. As it happens, I think it *is* possible to capture commonsense beliefs through experiment, as I will argue in this thesis.

The main purpose of this thesis is to show that Vihvelin's burden of proof argument fails in light of the available empirical data. I argue that Vihvelin's scepticism of philosophical intuitions about free will and determinism does not warrant scepticism of all empirical results on these matters. Rather, it tells us to be careful about how experiment materials are designed – a concern shared by researchers. With such considerations in place, I will argue that the available data strongly indicates that Vihvelin is wrong in her claims about commonsense beliefs. From this, I will demonstrate how Vihvelin's burden of proof argument fails. Finally, I will look at the implications this has on Vihvelin's wider argument for a "commonsense compatibilism" as well as the implications for free will burden of proof arguments more generally.

The structure of the thesis is as follows:

Chapter 2: Here I will carefully describe Vihvelin's argument for why the higher burden is on the incompatibilist. First, by explaining her proposed way of classifying the terms of the debate and how this affords different modal claims to incompatibilists and compatibilists. Second, by explaining how Vihvelin assesses burden of proof, appealing to both commonsense beliefs and the different modal claims made by each side of the debate.

Chapter 3: I will look closely at Vihvelin's claims about commonsense beliefs and show why these are empirical claims that can only be decisively resolved through empirical means. And I will also present the problem Vihvelin's scepticism of philosophical intuitions poses for empirical research on free will and argue that such research can nonetheless be successful. I will develop Vihvelin's claims about commonsense beliefs into three testable hypotheses:

H_1: Most people do not hold a belief about whether or not determinism is true.

H_2: Most people do not hold a belief about whether or not free will is compatible with determinism.

H_3: Most people believe that they have the ability to do otherwise, but do not believe this is an unconditional ability.

Chapter 4: Here I will examine the data from psychology and experimental philosophy of free will, demonstrating that the evidence indicates that all three of the hypotheses developed in Chapter 3 are false.

Chapter 5: At this point I finish the counterargument to Vihvelin's burden of proof argument by exposing how the evidence bears on it. First, by making clear how the evidence impacts claims about commonsense beliefs. Second, by reassessing the argumentative burdens, according to Vihvelin's method, but in light of the evidence.

Chapter 6: Finally, I will look at some wider implications of these results. First, the implications on Vihvelin's attempt to make a case for commonsense compatibilism. Second, the implications for claims about burden of proof in the free will debate more generally.

Chapter 2: Vihvelin's Burden of Proof Argument

A central aim of Kadri Vihvelin's (2013) *Causes, Laws, and Free Will* is to show that it is a mistake to think that determinism entails we never have the ability to do otherwise (p. 6). She wants to show that a conditional reading of an ability to do otherwise is coherent and is consistent with the commonsense view of free will (p. 6). On this basis she develops an argument for the compatibility of free will and determinism, and she calls this an argument for "commonsense compatibilism" (p. 33). As the name implies, commonsense compatibilism aims to be consistent with commonsense beliefs about free will as well as compatible with determinism.[1] This provides the groundwork for an account of commonsense compatibilism she calls the Bundle of Dispositions view (p. 169).

Vihvelin's case for commonsense compatibilism and the Bundle of Dispositions view depends on a number of claims she makes about what the commonsense beliefs about free will happen to be. Importantly, these claims are employed in an argument regarding who has the burden of proof in the debate between the compatibilist and the incompatibilist (p. 29). Vihvelin claims that the commonsense view is not incompatibilist, nor is it compatibilist, and that this means both sides of the debate are burdened with the need to make an argument (p. 30). There is another basis for assigning burden of proof though, she argues, which has to do with the different modal claims made by each side of the debate. She argues that the incompatibilist makes a more restrictive modal claim than the compatibilist and this results in the incompatibilist having a higher burden of proof than the compatibilist (p. 30). In this chapter I will go through her burden of proof argument in detail.

2.1 The Three-Fold Classification

To understand Vihvelin's burden of proof argument, it is necessary to understand the framework she uses for classifying the various sides of the free will debate. This will allow us to grasp how her definitions of "compatibilist" and "incompatibilist" differ from the standard definitions and to understand what she means by "impossibilist". Understanding this

---

[1] Vihvelin notes that her account is also compatible with indeterminism, hence her book's subtitle, "Why determinism doesn't matter". Her main challenge though is to show compatibility with determinism, and it is this claim I focus on for this thesis.

framework will also help to make sense of her appeal to the modal claims that she believes are made by the compatibilist and the incompatibilist. Mickelson (2015) has helpfully dubbed Vihvelin's classification the "Three-fold Classification", and I will refer to it as such.

2.1.1 A New Classification for an Old Debate

A central concern of free will debates concerns the logical compatibility of determinism and free will. In this context, "compatibilist" refers to someone who believes that free will and determinism are compatible, while "incompatibilist" refers to someone who believes they are incompatible. This often gets cached out in terms of moral responsibility, for example Pereboom (1995) says "compatibilism is the view that determinism is compatible with whatever sort of freedom is sufficient for moral responsibility, while incompatibilism is the view that determinism is not compatible with this sort of freedom" (p. 42). Vihvelin (2013) is very clear that her concern is *not* with moral responsibility, but with the metaphysical question of *free will's* compatibility with determinism (p. 20).[2]

Vihvelin argues that the contemporary debate around the compatibility of free will and determinism has lost some important features of traditional debates. Recovering these features is one of Vihvelin's motivations for introducing the Three-fold Classification. She describes the traditional problem of free will and determinism arising as an "*empirical worry*" in the face of science's success in explaining the natural world (p. 26). She says this worry was based on two assumptions: "i) that we have, or at least appear to have, free will; ii) if determinism turned out to be true, we would not have free will" (p. 26). This situated the debate, according to Vihvelin, as "between participants we would now call 'hard determinists' (the defenders of science) and the participants we would now call 'libertarians' (those who argued that even science must have some limits, and the limits must leave room for the existence and exercise of our free will)" (p. 26). Vihvelin says that later philosophers noted that both sides of the debate were making the same assumption: that free will is incompatible with determinism. If

---

[2] At least, this is her concern at the stage of the debate I will be looking at for this thesis. She is also very much interested in the compatibility of moral responsibility and determinism but makes clear that she takes there to be an important distinction between the two questions. It's also worth noting that the use of these terms in philosophy gets messy. "Compatibilist" is also often used as shorthand for someone who believes both the logic claim that free will and determinism are compatible *and* an actuality claim: that we do in fact have free will. Vihvelin refers to herself as a compatibilist in this sense, but her strict definition of the term, as we will see, only implies a belief in the possibility of free will, not necessarily its actuality. And there are other compatibility questions in the free will debate, such as the compatibility of free will and *indeterminism*, but "compatibilist" and "incompatibilist" are usually used to refer to positions on the compatibility of determinism and free will or of determinism and moral responsibility, rather than other compatibility questions. I will clarify exactly how I will use the terms once I have explained Vihvelin's definitions in the following section.

this assumption is mistaken, then there is a third position available: that of the compatibilist. At this point in the traditional debate, according to Vihvelin, all participants – hard determinists, libertarians, and compatibilists – shared a significant assumption: that free will is possible. Unlike the old debates, Vihvelin claims, the new debates include people who believe free will is *impossible*. She says this leads to a debate between those who believe free will is possible and those who believe it impossible. Vihvelin sees this debate as a distraction from the more important debate regarding whether or not free will is compatible with determinism.[3] One of Vihvelin's motivations for the Three-fold Classification is to draw the lines of the debate in a way that makes clear that the debate between the incompatibilist and the compatibilist is different from the debate between those she calls the "possibilist" and the "impossibilist" (p. 24).

2.1.2 Possibilists and Impossibilists

To understand exactly what Vihvelin means by "possibilist" and "impossibilist", we must note that she is employing a Lewis-Stalnaker possible worlds semantics. This means, in short, that something is possible just in case there is a possible world in which it is true. This is not the only notion of possibility employed in free will debates, so the fact that Vihvelin's Three-fold Classification requires this notion is one of the ways in which it differs from the standard classification.

With this notion in mind, Vihvelin (2013) distinguishes between two questions:

> 1. Is it possible that non-godlike creatures like us have free will?
> 2. Is it possible that non-godlike creatures like us have free will *and* that [determinism] is true? (p. 25).

The "possibilist" is someone who believes that there is at least one possible world in which non-godlike creatures like us have free will. The "impossibilist" believes there are no such possible worlds. So, the possibilist answers "yes" to the first question and the impossibilist answers "no". Vihvelin notes that how you answer the first question determines your available answers to the second question. If you answer "no" to the first, then your answer to the second

---

[3] Perhaps Vihvelin is simplifying the history of the debates for the sake of brevity, but there are at least a couple of concerns worth raising with her account. The term "incompatibilist" was first used in print by Keith Lehrer in 1960, with van Inwagen (1983) popularising the term. So, despite earlier philosophers arguing that free will and determinism were compatible, they did not characterise themselves as engaging in the debate in quite the way Vihvelin's account suggests. It is also not clear that 'hard determinists' really did all agree that free will was possible. Nevertheless, for the sake of explaining Vihvelin's classification of terms, I will let the problems with this potted history slide.

must also be "no".[4] On the other hand, if you answer "yes" to the first question, then you can answer "yes", "no", or "don't know" to the second. This means, according to Vihvelin, the problem of whether or not free will is compatible with determinism "arises *only* for those who answered 'yes' to the first question – the possibilists" (p. 25).[5] Vihvelin's Three-fold Classification attempts to capture this point by stipulating that "compatibilists" and "incompatibilists" both be deemed "possibilists" – they both believe there is at least one possible world in which creatures like us have free will. Under this delineation, compatibilists and incompatibilists are distinguished by their answer to the second question. The compatibilist answers "yes" – believing there is at least one possible world in which creatures like us have free will and where determinism is true. The incompatibilist answers "no" – believing there are no such possible worlds.

This brings out the important definitional difference between the standard classification of terms and the Three-fold Classification. Under the standard classification, the incompatibilist can *also* believe that free will is impossible. According to Vihvelin's classification, if you believe free will is impossible, for any reason, you cannot call yourself an incompatibilist. As she puts it, this proposal is "unorthodox" (p. 25).[6]

---

[4] As Vihvelin points out, if you replace "determinism" in the second question with any other claim (such as indeterminism, physicalism and so on), the impossibilist must always answer "no" (p25).

[5] This isn't quite right. Someone might answer "don't know" to the first question and also find the second question meaningful. I will return to this point soon.

[6] There are some significant problems with this way of defining terms that need to be acknowledged. One problem is that there is a third answer available for each question: "don't know". Suppose someone answers "don't know" to the first question and moves on to the second question precisely as a means to help determine an answer to the first question. Being unsure of whether we can possibly have free will, they ask if we can have it in a deterministic world. Reasoning that we can't, they then ask if we can have it in an indeterministic world. Reasoning that this is also impossible, and having exhausted the apparent causal options, they can return to the first question, answering "no". Importantly, under the standard definition of terms, this reasoner can claim they are an incompatibilist about free will and determinism (in part) because they answer "no" to the second question. Vihvelin recognises that some philosophers go through this process of reasoning but says that her classification is based on the *claims* people make, not their *reasons* for making such claims. As such, someone who comes to claim that free will is impossible is defined as an "impossibilist", regardless of the reasoning that got them there. But this appeal to claims over reasons doesn't really work. Remember that the first claim made in this reasoning process was that free will is incompatible with determinism. Why should this claim be ignored once the reasoner makes the further claim that free will is impossible? And having a claims-based classification risks obscuring important aspects of the reasoning people might go through. If someone can say that they believe free will is impossible because they are an incompatibilist then it is clear that a successful compatibilist argument could convince them that free will is possible. Vihvelin says that the debate between the compatibilist and the incompatibilist is a question that only arises for those who answer "yes" to the first question. Is this a stipulation or a logical consequence of her stipulated definitions? It seems to be the latter, but this would betray that this is not simply a claims-based classification. And in this case our imagined reasoner has *very good* reasons for taking part in this debate. Despite problems with Vihvelin's classification, I will put such worries aside so I can continue recounting her burden of proof argument. See Mickelson (2015) for further problems with the Three-fold Classification.

For this thesis, I will use "incompatibilist" and "compatibilist" to refer to Vihvelin's definition of the terms, with caveats regarding the empirical literature, as will be discussed in Chapter 4. That is, both terms refer to people who claim free will is possible, and their claims are about the compatibility of *free will* and determinism, not moral responsibility and determinism.

2.2 Two Sources of Burden

As we go through her argument, we will see that Vihvelin actually has two different sources for applying argumentative burdens. The first regards commonsense beliefs, taking it that those who argue against commonsense beliefs have the burden of proof. The second regards possibility claims, taking it that those who argue something is impossible have a higher burden of proof than those who argue that thing is possible. This gets tricky, so before I look at the burdens in the debate between the incompatibilist and the compatibilist it will be helpful to start with Vihvelin's first example – the debate between the impossibilist and the possibilist.

2.2.1 The Impossibilist's Burden of Proof

When discussing this example, Vihvelin begins by considering how the claims of each side of the debate between the impossibilist and the possibilist compare to commonsense beliefs:

> In the debate between the *impossibilist* and the possibilist, the burden of proof is on the impossibilist, for the impossibilist argues against a belief that is firmly embedded at the core of our commonsense beliefs about the world. We believe, and arguably cannot help believing, that we have and often act with free will. It does not follow that we *actually* have free will nor does it even follow that it is *possible* for us to have free will. (Since we are not infallible, it's possible for us to be mistaken in our beliefs about what's possible.) But in the absence of an *argument* that our commonsense belief that we have free will is or entails the belief that an impossible state of affairs obtains, we are entitled to believe that it is at least *possible* for us to have free will. (p. 28)

It is unclear *why* Vihvelin appeals to commonsense beliefs for her burden of proof argument, but presumably it is because she takes the debate to be about the concept FREE WILL and that this concept is reflected in commonsense beliefs. In this case, the commonsense belief is that we have free will and so we are entitled to believe we do have free will as long as there is no successful argument for free will being impossible. As such, the impossibilist has the burden of proof – they must make their case. The implication here is that the possibilist does not need

to make an argument since their view does not argue that a core commonsense belief is false (p. 28).[7]

When Vihvelin describes the next stage of the dialectic, we are introduced to the second source of burden – that regarding possibility claims:

> Once the impossibilist has produced an argument, the onus is on the possibilist to show what is wrong with the argument—that it is invalid, or has a false premise, or equivocates in some crucial respect, or is defective in any of the other ways in which arguments may be defective. The possibilist does not need to respond by providing a counterargument; in particular, the possibilist does not need to respond by providing an *analysis* of free will, or of acting with free will, or even a *sufficient condition* of free will or acting with free will. It is enough to describe cases of persons who have or act with free will with enough detail to make it plausible that these cases describe something that is really possible. (pp. 28-29)

Putting aside the curious use of "counterargument" here (isn't showing that something is "invalid, or has a false premise" etc providing a counterargument?), there is a subtle point being made that Vihvelin draws out more when describing the incompatibilist's burden. When someone argues that something is *impossible*, Vihvelin claims, they have a higher argumentative burden than someone who argues that thing is *possible*. While the person arguing that something is possible need only describe a case in sufficient detail as to make it seem possible, the person arguing that it is *impossible* must provide an argument for why such a case leads to a logical contradiction (p. 30).[8]

Vihvelin applies the two sources of burden, in order. First: those who argue against commonsense beliefs have the burden of proof. Second: those who argue something is impossible have a higher burden than those who argue it is possible. In this case, the impossibilist has the higher burden at both stages: their argument entails that a commonsense belief is false *and* they claim something is impossible.

2.2.2 The Incompatibilist's Burden

I now turn to how this approach to argumentative burdens impacts the nature of the debate between the compatibilist and the incompatibilist. As before, the first stage is to consider how

---

[7] This would imply there is a commonsense belief that free will is possible. But Vihvelin never outright says there is such a belief. Rather, she talks about the belief *in free will*, not in the *possibility* of free will. However, her argument seems to imply she takes there to be such a belief.

[8] This source of burden appears to be a consequence of her proposed classification of terms, but it's not clear to what extent it needs to be. Lycan (2003) makes a very similar argument for the burden of proof being on the incompatibilist due to the impossibility claim, but without the complications of a new classification.

each side of the debate compares to commonsense belief, assuming the burden of proof lies on those who argue against commonsense belief. In this case Vihvelin claims there is no commonsense belief in play. She asserts:

> common sense has either never considered, or never taken seriously, the thesis of determinism and because of this has *no view*, one way or the other, about the Determinism problem. (p. 29)

Three points of clarification are needed here. First, "the thesis of determinism" refers to the definition Vihvelin gives on the first page of her book:

> Determinism is the thesis that the laws of nature, together with the state of the universe at any time, entail the state of the universe at all later times. (p. 1)

Second, "the Determinism problem" is the problem of whether or not free will is compatible with the thesis of determinism.

Third, note that she uses "common sense" in the noun form rather than in the adjective form, such as when she talks of "commonsense belief". Vihvelin switches between these two forms liberally, but we should be careful in noting their differences. "Common sense" as a *noun* refers to an ability to make judgments (usually judgments of particular kind, e.g., good, wise, sensible).[9] "Commonsense", the *adjective*, is used to denote the products of those judgments, such as "commonsense beliefs", "commonsense views" and so on. As in the above quote, she often anthropomorphises the ability to make these judgments, referring to "common sense" as something that "considers" and "has views". For this thesis, the main question concerns whether people actually have the *beliefs* Vihvelin claims they have, not whether or not they have a particular ability to judge. For this reason, I will minimise my reference to "common sense", the ability, and talk mainly of "commonsense beliefs".

Note that Vihvelin refers to both the ability *and* its products in the quoted passage. She tells us that "common sense" has not considered the thesis of determinism and therefore has "*no view*" on the Determinism problem (p. 29). In other words, she claims that people have not applied their ability to make (good) judgments to the question of whether or not determinism is true and therefore they have no commonsense beliefs about whether or not free will is

---

[9] I use "ability" here to refer to the various ways philosophers have described the kind of thing common sense is. The most prominent of terms used was "faculty", famously by Reid (1823). But faculty psychology has long since fallen out of favour, so philosophers who like to think of common sense in this way tend to use different terms such as "ability", "capacity" and so on. It is not clear at all how changing the *term* resolves the underlying issues with faculty psychology, but I will leave that discussion for another time.

compatible with determinism.[10] For this reason, so her argument goes, people are neither naturally compatibilists *nor* naturally incompatibilists. What does this mean when assessing the first of Vihvelin's two sources of burden? Vihvelin puts it that:

> neither the incompatibilist nor the compatibilist has the burden of proof so far as the debate concerning the Determinism question is concerned. In the absence of argument, the default philosophical position is agnosticism, not compatibilism or incompatibilism. Each side has to make her case by argument. (p. 30)

I should note that Vihvelin is not consistent with her use of "burden". While here she says that *neither* side of the debate has the burden of proof, in a later chapter she seems to suggest they *both* have a burden of proof (for example, the compatibilist "discharges her burden of proof" (p. 167)). This latter use is more consistent with her point that both sides of the debate need to make an argument. What she seems to mean is that both sides have a burden of proof, but that appealing to commonsense belief makes no difference to *the level* of burden on each.

As it was with the debate between the impossibilist and the possibilist, appealing to commonsense belief is only the first stage in assessing burden of proof. The second stage is to compare the possibility claims made by each side. Remember that Vihvelin assumes that Lewis-Stalnaker possible worlds semantics are appropriate for framing debates about possibility. Vihvelin's delineation of terms has it that a compatibilist believes there are possible worlds in which both determinism is true and that non-godlike creatures like us have free will. The incompatibilist denies there are such possible worlds. Like the impossibilist, the incompatibilist is claiming something is impossible, and like the possibilist, the compatibilist is claiming that thing is possible. So just as the impossibilist had the higher argumentative burden than the possibilist, the incompatibilist has a higher burden than the compatibilist. Here Vihvelin gives more details on why burden is applied in this way:

> Conceivability is not proof of possibility, but it is the best evidence we can have of possibility, so all that the compatibilist has to offer, by way of *positive argument*, is a careful description of a case of a person who has and acts with free will, together with the additional claim that determinism is true and with enough detail to make it clear what this additional claim amounts to. (p. 30)

The argument required by the incompatibilist is more demanding, being:

---

[10] It's not totally clear that failing to consider the "thesis of determinism" means having no belief about whether or not determinism is true. I will discuss this in Chapter 3 and explain why I think this formulation gets at the point Vihvelin was intending to make.

to show that the addition of determinism to the description results in impossibility and, more generally, to provide some *argument* for the claim that it is impossible for *any* deterministic agent to have or act with free will. The strength of the case for incompatibilism rests on the strength of her arguments. If the arguments fail, we are entitled to believe that the compatibilist's answer to the Determinism question is the correct one. (p. 30)

So, this second source of burden places a higher burden of proof on the incompatibilist than on the compatibilist.[11]

---

[11] Again, I am extrapolating Vihvelin's intended meaning here. She does not actually say the incompatibilist has a "higher burden of proof", but she does use this phrase later, when talking about the debates concerning the possibility of moral responsibility and its compatibility with determinism (p. 48). I will discuss this in more detail in Chapter 5 when I look at how the empirical evidence bears on Vihvelin's burden of proof argument.

Chapter 3: The Need for an Empirical Approach

As we saw in the previous chapter, Vihvelin (2013) appeals to commonsense belief as the first source of argumentative burden. When it comes to the debate between the impossibilist and the possibilist, the commonsense belief appealed to is that we have free will. Since the impossibilist disagrees with a commonsense belief and the possibilist agrees with it, the impossibilist has the burden of proof. Regarding the debate between the incompatibilist and the compatibilist, Vihvelin claims that there is no commonsense belief to appeal to. As such, both sides of the debate have the burden of proof.

Obviously, these assessments of burden depend on what the commonsense beliefs actually are. This is clearly an empirical matter, yet Vihvelin gives no hint that she has a serious empirical basis for her claims.[12] Her claim that there is no commonsense belief about the compatibility of free will and determinism is derived from her claim that there is no commonsense belief about whether or not determinism is true. But again, she gives no account of a rigorous empirical basis for this claim. So, it seems quite straightforward that we should check the relevant empirical evidence in order to confirm or deny Vihvelin's claims. Experimental philosophers have been researching free will beliefs for nearly two decades now, so it makes sense to look at that research to check what the best evidence says on the matter. However, Vihvelin sets up a potential obstacle to this approach. She claims that the kinds of judgments people make about thought experiments should *not* be a basis for burden of proof claims.

Vihvelin calls these judgments "philosophical intuitions" (p. 29), arguing that these should not be considered reliable guides to truth because it is too easy for such intuitions to be influenced by "irrelevant factors" (p. 243). As we will see in Chapter 4, a number of the studies on free will beliefs have vignettes modelled on thought experiments, so it seems Vihvelin would reject the evidence from such studies. So, before we can look at the evidence about what people's free will beliefs are, we need to get clear on what the research is actually capturing. Does any of it get at the commonsense beliefs Vihvelin claims should be appealed to when

---

[12] As I will later note, she does suggest she has some empirical support when she refers to the views of her students. However, there is no indication that these views were surveyed systematically and no indication of wider research.

assessing burden of proof, or does it get at the philosophical intuitions she claims should *not* be appealed to? The purpose of this chapter is to make the case for why we *can* use experimental philosophy to get at the kinds of commonsense beliefs Vihvelin is concerned with.

3.1 Commonsense Beliefs Are an Empirical Matter

As mentioned in Chapter 2, Vihvelin claims that one of our commonsense beliefs is that we have free will. This isn't controversial of course, but Vihvelin also makes a series of claims about what people mean by free will, many of which are controversial. For example, she says "we want to maximize our potential, develop our abilities as best we can, take advantage of our opportunities, acquire the ability to do the right thing for the right reason, live the best life it is in our power to live" (p. 1). She may or may not be correct about any number of these claims, but most are irrelevant to the purposes of this thesis. The claims that matter here are those pertaining to beliefs about determinism and the compatibility of free will and determinism.

3.1.1 Vihvelin's Claims About Commonsense Beliefs

In Chapter 2 I put it that Vihvelin claims there are no commonsense beliefs about whether or not determinism is true or about whether or not free will is compatible with determinism. However, this is not Vihvelin's formulation. I believe this is what she intends, but her actual formulations will take some disambiguating to make this clear.

Here is exactly what she says:

> I believe that common sense has either never considered, or never taken seriously, the thesis of determinism and because of this has *no view*, one way or the other, about the Determinism problem. (p. 29)

> I think, as I said before, that common sense has no opinion, one way or the other, about the truth of determinism, and therefore has never had to confront the question of what would be the case if determinism were true. (p. 129)

Let's go through the various aspects of these statements step by step. As I noted in Chapter 2, Vihvelin switches between "common sense", the noun (referring to the ability to make good judgments), and "commonsense", the adjective (referring to the products of this ability). Though she explicitly refers only to the ability in the above formulations, she implicitly refers to its products. That is, "common sense has no opinion" implies "there is no commonsense opinion". And "*no view*" in the second formulation means "no commonsense view". Further,

we can translate words like "opinion" and "view" to "belief". So, the above passages are claims about commonsense beliefs.

Remember that Vihvelin defines "the thesis of determinism" as "the thesis that the laws of nature, together with the state of the universe at any time, entail the state of the universe at all later times" (p. 1). And the "Determinism problem" is the problem of whether or not free will is compatible with determinism (p. 24). But does Vihvelin really mean to say that ordinary people's views on the compatibility of free will and determinism depend on them having considered a specific, philosopher's definition of determinism? I don't think so. *Of course* it's true that most people have never considered the thesis of determinism, *so defined*. But this is *trivially* true. No-one claiming that people are naturally incompatibilists seriously thinks that this means ordinary people have heard of such a specific thesis – that's not what is at stake when assessing who has burden of proof. To make this clear, let's look at two examples of how proponents of natural incompatibilism have framed their view. First, Kane (1999):

> In my experience, most ordinary persons start out as natural incompatibilists. They believe there is some kind of conflict between freedom and determinism; and the idea that freedom and responsibility might be compatible with determinism looks to them at first like a "quagmire of evasion" (William James) or "a wretched subterfuge" (Immanuel Kant). (p. 218)

When Kane says "ordinary persons start out as natural incompatibilists" he surely doesn't mean that they have a pre-theoretic position that free will is incompatible with a philosopher's particular thesis of determinism. Note how Kane quickly shifts to the claim that ordinary people are unconvinced by compatibilist arguments. Obviously he doesn't mean that most ordinary people have *heard* compatibilist arguments. He is talking about how ordinary people respond to compatibilist arguments *upon* coming across them. Likewise, in saying that people are natural incompatibilists Kane just means that people are naturally inclined to take incompatibilist positions upon being confronted with the problem.

For a less ambiguous view, here is Nichols (2015) talking about the empirical evidence for incompatibilist intuitions:

> It's surprising that people have intuitions about this at all. Many, perhaps most, of these participants have never heard of causal determinism before being exposed to the vignettes. And yet people converge—even across cultures—on the view that determinism precludes moral responsibility. The flat-footed interpretation of the results is that people really do have some commitment to incompatibilism, and describing determinism to them affords them the opportunity to express that commitment. I think this flat-footed interpretation is the right one. (p. 77)

Nichols explicitly says that most ordinary people have probably never heard of the philosopher's concept of determinism, but that they are nevertheless committed to incompatibilism.

Since Vihvelin (2013) is clearly taking part in the same debate as that of Kane and Nichols, I do not take her to be making the trivially true claim. The claim I think she *is* making is better captured in her second formulation: "common sense has no opinion, one way or the other, about the truth of determinism" (p. 129). I take this to mean: "there is no commonsense belief about whether or not determinism (in a suitably vague sense) is true". This is a more serious question because it leaves the definition of determinism vague enough to be filled in with whatever concept the folk happen to have.

Next let's look at what we should make of the alternatives given in the first formulation: "common sense has either never considered, or never taken seriously, the thesis of determinism and because of this has *no view*, one way or the other, about the Determinism problem" (p. 29). Here it seems Vihvelin is giving two alternative reasons for thinking that there is no commonsense belief about whether or not determinism is compatible with free will. Either people have never considered determinism, in which case they wouldn't have a view whether or not it was compatible with anything. Or they have considered it, but not taken it seriously, again leaving them without a view on its implications. With this I am ready to reformulate these two statements.

I take it that the first says: "Ordinary people have either never considered determinism or never taken it seriously, therefore there is no commonsense belief about whether or not free will and determinism are compatible".

I take it that the second says "There is no commonsense belief about whether or not determinism is true, therefore there is no commonsense belief about whether or not free will and determinism are compatible".

So, they have different premises for the same conclusion. However, I think the premise of the second formulation is a tacit premise in the first one. In this case the reasoning runs "there is no commonsense belief about whether or not determinism is true, nor has determinism been considered (at least not *seriously*), therefore there is no commonsense belief about whether or not free will and determinism are compatible."

With that, we are one step away from making clear the specific claims Vihvelin is making about folk belief. Though Vihvelin never makes this explicit, I think we can take her use of "commonsense belief" to amount to, "a belief shared by most people". This gives us three unambiguous claims about folk belief:

Most people do not hold a belief about whether or not determinism is true.

Most people have never seriously considered determinism.

Most people do not hold a belief about whether or not free will is compatible with determinism.

So, we have three negative claims about folk beliefs. Now let's look at closely related positive claim. Vihvelin says:

the belief that we have the ability to do otherwise is conceptually central to our commonsense view of ourselves as free and responsible agents. (p. 6)

At this point I need to point out that Vihvelin favours a conditional analysis of "ability to do otherwise" rather than an unconditional analysis. On a *conditional* analysis, the phrase becomes a conditional statement about what *would* have happened, had things been different, e.g., "if the past (or the laws of physics) had been different then I would have done otherwise". On an *unconditional* analysis, the phrase is read as referring to the actual circumstances, e.g., "I was able to do otherwise even if everything had been exactly the same right up until the moment of choice".[13] Philosophers who favour the conditional analysis do so to make "ability to do otherwise" (or "could have done otherwise") compatible with determinism. It is argued that determinism says nothing about the initial conditions of the universe, so it would not preclude the possibility that those conditions could have been different, and if they *were* different then it's possible that we would have made different decisions. The unconditional analysis is taken to preclude determinism. If determinism is true, then the state of the universe and the laws of physics fully determine all other states of the universe. So, if "ability to do otherwise" means "ability to do otherwise in exactly these conditions" then this would only be possible if determinism was false.

---

[13] Vihvelin's conditional analysis is rather more complicated than this. She takes "ability" to be a matter of dispositions, and favours a modified version of Lewis's Reformed Conditional Analysis (p. 170). As we will see though, the specific analysis she favours is not relevant to the questions I want to answer.

Though Vihvelin favours a conditional analysis, she does not claim that the commonsense belief of an ability to do otherwise is conditional. It is unclear, but I think she can be read as meaning that ordinary people have neither a conditional nor an unconditional concept of this ability. Regardless, Vihvelin's argument requires the commonsense belief to be *consistent* with a conditional analysis. If people actually have an *unconditional* concept of the ability to do otherwise, then it would be much harder for Vihvelin to argue that there is no commonsense beliefs about determinism or the compatibility of free will and determinism. Believing that we can only do otherwise in the exact conditions we find ourselves in seems to imply a belief in both indeterminism and a belief that determinism would prevent us from having free will.[14] So, Vihvelin is at least claiming that the commonsense belief in the ability to do otherwise is *not* of an unconditional ability. Taking "commonsense belief" to refer to "a belief shared by most people", as we did earlier, this can be translated into the claim:

> Most people believe that they have the ability to do otherwise, but do not believe this is an unconditional ability.

3.1.2 Why Should We Believe These Claims?

At this point I have listed four unambiguous claims, all of which are clearly empirical claims. So, on what basis should we believe Vihvelin is right about these claims? Looking back at the earlier quotes, she says, "I believe common sense has…" and "I think…common sense has…". Ok, but why should we think that what Kadri Vihvelin happens to believe about commonsense beliefs is actually true? She does give us a hint about why she believes these things when talking about how people purportedly respond to the Manipulation Argument:

> I am confident, on the basis of my experience of years of teaching the free will/determinism problem to students at all levels, that not everyone would respond in this way. (p. 154)

Vihvelin also talks about how her students respond to deterministic scenarios more generally (p. 143). So perhaps Vihvelin believes these things because of her experience with her students. But she gives no indication that she has done any systematic research of her students' beliefs. And, even if she had, why should we think that students studying philosophy in the particular universities Vihvelin has taught at are representative of the general population? Either way, it seems quite straightforward that systematic research would be the best way to confirm or deny Vihvelin's claims about commonsense beliefs. If such research confirms these claims, then

---

[14] Though it need not imply this (folk beliefs might be inconsistent).

what she says about burden of proof has a solid foundation. If not, then those foundations look weak. I will talk more about the potential implications of empirical results in Section 3.3. Before I can do that though, it's necessary to examine some statements suggesting Vihvelin may not actually be convinced by the empirical data.

3.2 The Problem Posed for Empirical Research

When discussing common assumptions about who has the burden of proof in the debate between the compatibilist and incompatibilist, Vihvelin says:

> I do not deny that philosophical intuition (which I define as the intuitions produced by exposure to philosophy, whether in the form of thought experiments or definitions of unfamiliar terms like 'determinism') supports incompatibilism against compatibilism. However, I deny that this gives us grounds for saying that the burden of proof is on the compatibilist. (p. 29)

She says it is a mistake to think the burden of proof should always be on those who argue against philosophical intuition, explaining:

> intuitions have evidential weight only insofar as we have reason to believe that they are reliable guides to truth. This is an empirical question, and while some intuitions may be highly reliable guides to truth, I don't know any good reason to believe that philosophical intuitions, *in general*, and intuitions about free will and determinism more specifically, are reliable guides to the truth. (p. 29)

It is unclear exactly what we should take from this since Vihvelin does not make these remarks with direct reference to empirical research on free will beliefs, but this seems to pose a problem for such research in the following way.

Vihvelin defines "philosophical intuitions" as those intuitions produced by "exposure to philosophy" (p. 29). Here she hasn't told us what she means by "intuition", but elsewhere she says, "intuitions are just a kind of belief" (Vihvelin, 2015, p. 399). Let's unpack why the claim that such intuitions should not be appealed to when assessing burden of proof is a problem for empirical research.

Suppose a study has participants read a thought experiment before asking them for their judgments about the scenario described in the thought experiment – as a number of studies do, (such as classic examples, Nichols & Knobe, 2007; Nahmias E. , Morris, Nadelhoffer, & Turner, 2005). According to Vihvelin, their responses should be classed as "philosophical intuitions". Since she takes such intuitions to be unreliable guides to truth, they would give us no good reason to place a burden of proof on those who argue against them. If this is right, then

we can't simply look at the data from such a study and say that we have discovered what the commonsense beliefs are and where to apply the burden of proof. If we did try to apply a burden on this basis, it seems that Vihvelin would reject it, complaining that we haven't necessarily captured the commonsense beliefs at all.

This is not to say that Vihvelin thinks a philosophical intuition can never be identical to a commonsense belief. Let's make this clear with a simple example. Suppose a study uses a thought experiment to test if participants believe they have the ability to do otherwise and most respond that they do. This is consistent with one of Vihvelin's claims about commonsense beliefs. So, she might say that even though participants are having philosophical intuitions (they expressed an intuition that comes from being exposed to philosophy), they also happen to be the same as the commonsense belief. What is unclear is whether or not Vihvelin would think such an experiment actually gets at the commonsense belief or if she would just take it as a coincidence that elicited intuitions happen to be identical with a commonsense belief. If it is the former then we might be able to figure out how an experiment can use thought experiments to get at commonsense beliefs, despite Vihvelin's concerns with philosophical intuitions. If it is the latter, then it's not clear how a study using thought experiments could be convincing to Vihvelin.

Now, I take Vihvelin's characterisation of commonsense beliefs to roughly mean "aspects of folk concepts". If this is right, then the collection of commonsense beliefs about free will that Vihvelin lists throughout her book are claims about what the folk concept FREE WILL is made up of. With that in mind, let's distinguish between two ways in which we can talk about philosophical intuitions being "unreliable". First, there is the question of whether or not philosophical intuitions are reliable guides to truths *about* concepts, e.g., can philosophical intuitions tell us whether there is some aspect of FREE WILL that implies it is incompatible with determinism? Second, there is the question of whether or not philosophical intuitions can reliably tell us what the concepts *are* (what the commonsense beliefs are). I take it that Vihvelin's scepticism about philosophical intuitions is concerned with the first question. But this scepticism suggests she also has reason to be concerned with the second question.

Now, an experiment that asks participants about the possibility of free will existing in a deterministic scenario tries to get at what the concept *is* via testing what people take its *implications* to be. So, the problem posed is this: maybe empirical research that exposes participants to philosophy could cause people to have intuitions about the implications of their

concept FREE WILL that *do not* actually follow. So, though Vihvelin does not express this worry, I think that her reasons for thinking philosophical intuitions are unreliable guides to truth could also give her reason to worry that much of the empirical research fails to capture commonsense beliefs.

It's unclear how far this problem might go. As mentioned above, a number of well-used vignettes could arguably be described as thought experiments, e.g., (Nichols & Knobe, 2007; Nahmias E. , Morris, Nadelhoffer, & Turner, 2005). Would Vihvelin discount the evidence from all studies utilising vignettes? There are a number of studies that avoid using items that can be reasonably described as thought experiments, e.g., (Nadelhoffer, Shepard, Nahmias, Sripada, & Ross, 2014; Paulhus & Carey, 2011). But would these avoid Vihvelin's scepticism, or could they too be problematic in some way?

To get a better handle on this, I will look more closely at what Vihvelin thinks makes philosophical intuitions unreliable guides to truth.

3.2.1 The Reliability of Philosophical Intuitions

Unfortunately, Vihvelin (2013) does not say a lot about why she thinks philosophical intuitions are unreliable, but we can glean something helpful from a footnote:

> Such intuitions are all too easily caused by irrelevant factors—for instance, the kinds of factors that are exploited in arguments for *logical fatalism*. (p. 243)

This is helpful because it means she is concerned with how thought experiments (or definitions of terms) might influence people to think things that are not correct about the subject matter. If participants in a study are misinterpreting the scenarios and questions being asked, then their answers may not reflect their actual beliefs about the topic of the experiment.

Vihvelin's (2013) example regards fatalism. Vihvelin talks more about conflating determinism with fatalism in the short paper *How Not To Think about Free Will* (2015):

> It's well known, in philosophy, that the fatalist is confused. Truth isn't the same as necessity, of any kind. The fact that there are truths about my future choices and actions does not affect my freedom *in any way*. But many years of trying to explain to my students why the fatalist is confused has convinced me that fatalist thinking runs deep. Some students get it; others never do…Even though it's a mistake, many people have the intuition that if it is "already true" what our future will be, then our future is not up to us; they think that truth alone – *regardless of determinism* – would rob us of free will. But if determinism is true, then there are detailed and specific truths about *all* our future choices and actions. So the intuition that determinism robs us of free will should not be trusted, for it might be a fatalist intuition in disguise. (p. 400)

Here we can imagine what Vihvelin (2013) might mean by "irrelevant factors". If a thought experiment that is supposed to describe determinism causes people to respond to it with judgments that are not about the implications of *determinism* but about the implications of *there being truths about our future decisions,* then there is a risk that they make the fatalist fallacy. The thought experiment may have caused people to believe something that is inconsistent with the intended scenario, causing an intuition that "should not be trusted". This helps explain why Vihvelin takes thought experiments to result in intuitions that are unreliable guides to truth. If philosophical intuitions are often the result of logical fallacies, then they would not be reliable guides to what is true. It also gives us a specific reason to worry that some empirical research could fail to capture commonsense beliefs. For example, if a study design causes participants to conflate a deterministic scenario with a fatalist scenario, then their responses may not tell us what they believe about the implications of determinism.

Vihvelin (2013) also notes she is indebted to Talbot's doctoral dissertation for her views on the reliability of philosophical intuitions (p. 243). Talbot (2009) details a number of cognitive biases that should make us worry about the reliability of philosophical intuitions. For example, context effects such as the *order effect* and the *compatibility effect*, can lead to faulty interpretations of scenarios (p. 104). These are exactly the kinds of things empirical researchers concern themselves with when assessing experiment design. So, again, just as these potential problems should make us sceptical of the reliability of philosophical intuitions as guides to truth, we should also worry that poor experiment design could prevent us from capturing what participants believe about the subject under question.

Vihvelin (2015) quotes a vignette from Nichols and Knobe (2007), highlighting phrases that she thinks should make us question whether study participants would all be thinking about the same thing. I quote the vignette in full including Vihvelin's bolding of the questionable phrases:

> Imagine a universe (Universe A) in which everything that happens is **completely caused** by whatever happened before it. This is true from the very beginning of the universe, so what happened in the beginning of the universe caused what happened next, and so on right up until the present. For example one day John decided to have French Fries at lunch. Like everything else, this decision was completely caused by what happened before it. So, **if everything in this universe was exactly the same up until John made his decision**, then it *had to happen* that John would decide to have French Fries.
>
> Now imagine a universe (Universe B) in which *almost* everything that happens is **completely caused** by whatever happened before it. The one exception is human

decision making. For example, one day Mary decided to have French Fries at lunch. Since a person's decision in this universe is not completely caused by what happened before it, **even if everything in the universe was exactly the same up until Mary made her decision**, it **did not have to happen** that Mary would decide to have French Fries. She **could have decided to have something different**.

The key difference, then, is that in Universe A every decision is **completely caused** by what happened before the decision—**given the past, each decision *has to happen* the way that it does**. By contrast, in Universe B, decisions are not completely caused by the past, and each human decision ***does not have to happen* the way that it does**.

Which of these universes do you think is most like ours? (circle one) (2015, p. 398)

Vihvelin says that it is not at all clear what question is being asked here. She questions whether "completely caused" means the same as "had to happen, given the past," and "could not have decided anything different" (p. 398). Her suggestion, as far as I can tell, is that study participants might equate "completely caused" and "given the past, each decision has to happen the way that it does" with a lack of ability to do otherwise. Remember that Vihvelin (2013) believes the ability to do otherwise is a central commonsense belief (p. 6). She also takes this belief to be consistent with a conditional analysis of ability to do otherwise. So, if participants think that it's not possible to do otherwise in Universe A – even if past conditions had been different – then they have misunderstood the scenario. And if it really is a commonsense belief that we have the ability to do otherwise, then they would say Universe B is most like the actual universe, because that's the one in which they think we could have that ability.

This would be a problem because Nichols and Knobe's (2007) Universe A is meant to describe a *deterministic* universe, but not *necessarily* a universe without free will (p. 667).[15] The point of asking participants which universe is most like ours is to find out if they believe determinism is true, *not* if they believe in free will.

Now, I might be misunderstanding what Vihvelin (2015) sees as the problem, but it gets at her broad concern that participants might misinterpret thought experiments. Again, this helps explain Vihvelin's (2013) concern about philosophical intuitions being unreliable guides to truth. And it gives us reason to worry about whether studies like this capture commonsense beliefs.

The Nichols and Knobe (2007) study actually had 90% of respondents answer that Universe B was most like the actual universe (p. 669). This is taken as evidence that most

---

[15] Indeed, the study is not explicitly about free will at all, but *moral responsibility*. I will discuss the significance of this when I return to the study in Chapter 4.

people believe determinism is false. But if people are actually responding that way because they've misunderstood the thought experiment then we can't rely on that data to tell us just how many people really believe determinism is false.

We can see now that the problem posed for empirical research is that there is risk that experiment design may fail to capture commonsense beliefs because aspects of the experiment cause participants to have intuitions that do not reflect their beliefs on the subject being tested. In the following section I will look at how we can take this concern on board when looking at the existing empirical data.

3.3 On Examining the Empirical Data

Now, I don't want to put words in Vihvelin's mouth. She does not say that empirical research cannot get at commonsense beliefs. Rather, I have extracted a scepticism of the results that I think she might have. Vihvelin (2015) says of philosophising about free will: "Unless you know exactly what you are doing, and are sure you can do it well, avoid thought experiments (and avoid experimental philosophy)" (p. 398). So, she does not rule out the possibility of successful experimental work but warns that it is difficult. However, her comments do suggest a view that experimental work on beliefs about determinism and its compatibility with free will in particular is simply *too* difficult. Again, perhaps I am reading her wrong. But if this is her view, it betrays a dilemma. Either we can use experiments to answer these empirical questions, or we can't. If we can, then obviously we should look at the best available evidence (or conduct our own experiments). But if we *can't* get at empirical questions through experiment then our only recourses are silence or speculation. If we remain silent then we admit that we don't (and can't) know what the commonsense beliefs are, so there is no way to judge who has the burden of proof on that basis. And even if we speculate about commonsense beliefs, we cannot establish burden of proof on that basis. Admitting that it is mere speculation is to admit that we don't (and can't) know who has the burden of proof.[16] So either the experimental work can be successful, in which case the burden of proof debate should be settled with empirical data, or it cannot, in which case there is no burden of proof debate to be had. Thankfully, there are good reasons to think the empirical work *can* be successful, which I will look at now.

---

[16] If Vihvelin really does think it is impossible to get at commonsense beliefs about free will through experiment, then she should embrace this result. She would end up with exactly the same conclusion, the argument going: we don't know and can't know what the commonsense beliefs are so neither side has the burden of proof on that basis. But on the basis of possibility claims, the higher burden is on the incompatibilist.

3.3.1 How the Empirical Data May Succeed in Capturing Commonsense Beliefs

The problem posed in Section 3.2 is one that is essentially about what makes an experimental instrument successful at producing the data it is designed to probe. This is a central concern to all empirical researchers. In this case, the challenge Vihvelin's scepticism of philosophical intuitions brings is the potential for research that exposes participants to philosophy to be particularly prone to misinterpretation. Nonetheless, this challenge does not rule out the possibility that at least *some* experiments that expose participants to philosophy may accurately tell us what people believed prior to that exposure. The question is: which experiments do this most successfully?

When we look at the existing empirical research in Chapter 4, we will see that researchers have spent a lot of time trying to assess and improve experiment design. They have been concerned with whether or not their experiments really capture the beliefs they are intended to capture. This often includes looking at a number of the issues raised in the previous section. For example, order effects are often tested by randomising the ordering of questions and scenarios. And there has been a great deal of discussion about how vignettes and other instruments may be misinterpreted by participants. Such discussion often leads to new hypotheses, which can be tested in further experiments. These subsequent experiments are discussed further, leading to further hypotheses and so on.

Experimental work on beliefs about free will and determinism has been underway for decades now, with hundreds of papers published. The bulk of this work was done over the last 15 years in one of the main lines of research undertaken by the experimental philosophy movement. Researchers with a background primarily in psychology have often been concerned with matters that do not get at what philosophers of free will have been concerned with. And sometimes such researchers make philosophical errors. So, bringing in philosophical expertise has been helpful in improving experiment design.

So, despite the possibility that any one experiment may fail to capture commonsense beliefs, I have confidence that the existing data, when examined carefully, can tell us something useful about what it is people actually believe.

3.3.2 How to Approach the Evidence

When we look at the research, we need to have a clear idea of what we should expect to see if Vihvelin (2013) is right about commonsense beliefs and what we should expect to see if she is

wrong. I will now go through her main claims, formulated into the hypotheses listed in Section 3.1.1.

First, it is necessary to put aside one of the hypotheses as, unfortunately, there is little experimental work that can gives us clear answers on it. This is the hypothesis that most people have never considered determinism. Though I think this is the one claim Vihvelin gets right, we cannot strongly draw any conclusions about it from the data. But there are many studies on the other three hypotheses so let's go through those now, numbering them accordingly.

H$_1$: Most people do not hold a belief about whether or not determinism is true.

H$_2$: Most people do not hold a belief about whether or not free will is compatible with determinism.

H$_3$: Most people believe that they have the ability to do otherwise, but do not believe this is an unconditional ability.

What should we expect if the first hypothesis is correct? If it was true that most people had no belief about whether or not determinism is true, then we should probably expect them to answer questions on this with "don't know" or "neither agree nor disagree" (depending on the structure of the question). Or, if they are only given a binary option like "yes" or "no", then we should expect results to be indistinguishable from chance. Inversely, evidence against the first hypothesis would be a majority of participants expressing a position. Note that there are three ways this majority might break down. A majority believing in determinism, majority believing in indeterminism, or a majority made up of a combination of the two. That is, even if neither determinism nor indeterminism is the dominant belief, as long as most people have a view about the matter, it would be evidence against Vihvelin's claim.

I will not take such results as conclusive though. I will look for weaknesses in the experiment materials that may have produced results that do not reflect the beliefs participants held prior to the experiment. As we will see, a lot of that speculation has already been done, as have subsequent experiments to test these potential issues. So, I will also look at that further evidence.

For the second hypothesis, that most people are neither compatibilists nor incompatibilists, we should expect similar results as for the first. That is, evidence in favour of this claim would be participants mostly answering "don't know" or "neither agree nor disagree" to the relevant questions. Or, if they are only given a binary option, we would expect results

indistinguishable from chance. And inversely, evidence against this claim would be most participants responding with a view. Again, neither side needs to make up the majority on its own. The hypothesis looks false as long as most participants say they have a view at all.

As for the first hypothesis I will look for reasons why the experiments may fail to get at the beliefs people held prior to the study, and the subsequent research on that issue.

Finally, I will look at evidence for or against the third hypothesis, that most people believe in an ability to do otherwise and that this is not an unconditional ability. This hypothesis is a conjunction of two claims, and it is important that the evidence supports both for Vihvelin's claim about commonsense beliefs to stand. Evidence for this hypothesis will be most participants answering in agreement with questions or statements regarding the ability to do otherwise, but in disagreement with those that specifically test if this is an unconditional ability. Evidence against will be a majority agreeing that we have an unconditional ability to do otherwise. As for the other claims, I will examine whether evidence for or against the hypothesis stands up to scrutiny of the experiment materials.

This scrutiny of experiment materials is important for all the hypotheses. Though we don't know exactly how sceptical of the results Vihvelin may be, she is *right* to be concerned with how experiment materials might be interpreted. As we will see, researchers, and philosophers following the research, have raised many of the kinds of concerns we might expect Vihvelin to raise. And this has led to hypotheses for explaining away various results. Many of those hypotheses then get tested, sometimes producing supporting evidence, other times not.

But, after careful consideration of experiment materials and subsequent research, if the data still goes against any of these hypotheses, then this should be regarded as strong evidence against crucial part(s) of Vihvelin's burden of proof argument. The consequences the evidence has for Vihvelin's burden of proof argument will be looked at in Chapter 5.

Chapter 4: Surveying the Evidence

The earliest attempt at experimental work on free will and determinism was conducted by Nettler (1959), in which he purported to show that free will believers are more punitive than determinists. Pitting free will believers against determinists as if it is impossible to believe in both was only one problem with the study. In a misguided hostility toward philosophical definitions, Nettler's "determinism scale" shows no indication that he understood what philosophers have actually meant by the term (p. 379). A later study by Viney, Waldman, and Barchilon (1982), while making significant improvements on Nettler, also treated free will and determinism as mutually exclusive. Stroessner and Green (1990) were the first to correct this error, and also included two types of deterministic belief: religious-philosophical and psychosocial, finding most people believe in both. However, many of their items were of poor quality. For example, their "Psychosocial determinism" factor included: "My behaviours are limited by my background", "My exercise of free will is limited by my upbringing", "My free will is limited by such social conditions as wealth, career, and class" (p. 794). Of course, it is entirely possible for our choices to be *limited*, even if our social circumstances fail to fully determine our psychology and choices. Further, the kind of determinism philosophers have in mind in the current debates is scientific determinism: that the laws of physics and the initial conditions of the universe determine all subsequent events. So, agreement with such statements is no indication of how many people believe in the kind of determinism Vihvelin (2013) is concerned with. An unpublished scale (the FAD-4) developed by Paulhus and Margesson (1994) made further improvements, though still failed to make as careful distinctions as would be necessary for addressing the question at hand.

It's clear that early researchers either did not have the philosophical expertise to make the distinctions we are interested in, or they weren't interested in the same questions. Research coming out of the experimental philosophy movement over the last two decades has been more successful at getting at the questions we are interested in, so I will focus on this research here.

4.1 Experimental Philosophy on Beliefs About Determinism

In Chapter 3 I outlined what kind of evidence would support Vihvelin's claims and what would be evidence for her claims being wrong. Regarding beliefs about determinism, Vihvelin claims:

There is no commonsense belief about whether or not determinism is true.

As argued in Chapter 3 Section 3.11 we can take "commonsense belief" to mean "a belief shared by most people". So, the hypothesis we are testing here is:

H_1: Most people do not hold a belief about whether or not determinism is true.

What we should expect if this hypothesis is correct depends on the structure of the study. If participants are asked about their agreement with statements or scenarios describing determinism, then we would expect the mean of their answers to be close to the mid-point of the scale. For example, if using a Likert scale from 1-(strong disagreement) to 7-(strong agreement) we would expect the mean of answers to be not significantly different from 4-(neither agree nor disagree). If participants are given a binary choice, for example "yes" or "no" then we would expect results not significantly different from 50/50.

If the hypothesis is false, we would expect answers to reflect that most people have a believe about whether or not determinism is true. So, we would expect the mean on a Likert scale to be significantly different from the mid-point. If the choice is binary, we should expect results to be significantly different from 50/50. Note that it does not matter whether or not the majority are determinists, indeterminists, or some combination, Vihvelin's claim is that there is no belief either way. So long as there is evidence for most people holding a belief, this is evidence against her claim.

I also noted that there may be reasons to query results. The concern is that the results may not reflect participants' actual beliefs. For example, experiment design may cause participants to interpret deterministic scenarios as indeterministic. So, I will look at reasons to doubt the results. As we will see there has already been a great deal of discussion around these kinds of issues. And this has led to new hypotheses and further research testing those hypotheses. With that in mind, let's look at the research.

4.1.1 Studies on Belief in Determinism

The first study I will look at comes from Nahmias (2006). 99 participants read three scenarios describing a universe that is "re-created over and over again, starting from the exact same initial conditions (and with all the same laws of nature)" (p. 219). The scenarios were a lightning strike, choosing an ice-cream flavour, and choosing to steal a necklace. In each case participants were asked if the particular event described would "happen the exact same way" every time the universe is re-created. 30% gave the determinist response to all three scenarios,

40% gave the indeterminist response to all three, and 30% did not respond the same way for each question.

The low level of people responding "I don't know" provides evidence against Vihvelin's claim that people have no view about whether or not determinism is true. Could people have been influenced to answer "yes" or "no" despite not having any view about determinism prior to the experiment? Possibly. Perhaps the fact that 30% did not give the same answers for each question might indicate confusion about the scenarios. However, there are coherent ways to vary responses, such as believing that human choices are indeterministic while everything else is deterministic. Participants were given the opportunity to explain their answers, with 20% referencing something that might differentiate humans from natural processes like lightning strikes. However, only 9% answered in a way consistent with the view that decision-making was indeterministic and other processes deterministic.

As noted in Chapter 3, Nichols and Knobe (2007) found that 90% of participants believed an indeterministic universe was more similar to our own universe than a deterministic one. Participants were given a binary choice, so results not significantly different from chance would support the hypothesis that most people do not have a belief about whether or not determinism is true. That 90% expressed a belief in indeterminism appears to be striking evidence against this hypothesis.

However, as noted in Chapter 3, Vihvelin (2015) questions whether "completely caused", "had to happen, given the past", "could not have decided anything different" all mean the same thing, and whether participants may have a range of interpretations (p. 398). Her critique is vague, but it seems that her concern is with participants interpreting the "deterministic" universe as a world without free will because they take it that John (the character in the "deterministic" universe) is unable to do otherwise. This is actually the critique Nahmias (2006) makes, pointing out that determinism does not entail that we cannot do otherwise as long as you take into account the possibility of a conditional analysis (p. 222). Now, the vignette is designed to avoid this issue, as it says John's decision had to happen *if* everything was the same up until that point. However, it's possible that participants may fail to fully grasp the intended meaning here. I think the misinterpretation Vihvelin and Nahmias suggest is quite possible.

Another concern is that participants were given a binary choice. Perhaps there would have been a less striking result if people were asked their level of agreement on a Likert scale.

There is a general concern that applies to both the Nichols and Knobe (2007) study as well as the Nahmias (2006) study we looked at earlier. This is that the participants were all undergraduates at U.S. universities. These may not be reflective of wider populations such as the general U.S. population or the global population.

Given this limitation, Sarkissian, Chatterjee, De Brigard, Knobe, Nichols and Sirker (2010) conducted a cross-cultural study with participants across four countries using the Nichols and Knobe (2007) vignettes. The majority of participants in each case picked the indeterministic universe: United States 82%, India 85%, Hong Kong 65%, and Colombia 77% (p. 352). On the assumption that there is no systematic misinterpretation of the Nichols and Knobe vignettes, these are not good results for Vihvelin. It suggests that indeterministic views are even more widespread than previous studies indicated. A limitation of this research is that all participants did have at least one thing in common – all were undergraduate students at universities. Further the U.S., Hong Kong, and India participants were all English-speaking, though the Colombian participants read vignettes translated into Spanish (p. 352).

I now turn to a number of studies using single-sentence statements rather than the long vignettes used in the studies just mentioned. These are all attempts at designing a scale for measuring a variety of beliefs about free will. Nettler's (1959) "determinism scale" was the first of these. As noted earlier, Nettler was not concerned with measuring belief in the kind of determinism at stake in the free will debate. More recent scales, such as the FAD-4 (Paulhus & Margesson, 1994) and FAD-Plus (Paulhus & Carey, 2011), though superior to Nettler's scale, also failed to take seriously what philosophers of free will mean by "determinism". For example, the FAD-Plus, "Scientific Determinism" subscale had this item, "People's biological makeup influences their talents and personality" (p. 98) which is totally consistent with indeterminism.

Deery, Davis and Carey (2015) developed a superior scale, though not without issues. They created two determinism subscales to test for differences in belief in "proximal determinism" compared to "distal determinism" (pp. 795-796). The distal determinism items referred to how decisions were caused by events in the distant past – "the state of the universe millions of years ago" (p. 796). The proximal items referred to events in the immediate past, such as "neural processes" and "thoughts and desires" (p. 795). Participants were asked for their level of agreement with each item on a 7-point scale from 1 (strong disagreement) to 7 (strong agreement).

Deery et al. claim that both subscales are designed to measure belief in determinism because "it makes no philosophical difference whether the events that determine an action occurred recently, or a long time ago" (p. 790). So, if Vihvelin's claim that people have no view is correct then we should expect the means for each subscale to be not significantly different from the mid-point of 4 (neither agree nor disagree). The mean for the distal determinism items was 3.08, which is significantly different from the mid-point. However, the result for the proximal determinism was 4.06, which is *not* significantly different from the mid-point (p. 783). What should we make of these conflicting results? If Deery et al. are right that the subscales measure the same thing, then we should expect results that aren't significantly different from each other. But this isn't the case. So which descriptions (if either) are properly getting at participants' actual beliefs about determinism?

Morris (2015) makes a good argument for the distal determinism subscale being the more accurate measure. He points out the distal determinism subscale seems a stronger basis for inferring "determinism" as it is commonly described in the literature. He puts this common description as follows:

> *Given the state of affairs at any time t1, one could, in principle, infer on the basis of this state of affairs—and the laws of nature—the state of affairs at any subsequent time t1+n.* (p. 805).

Given the differing results, it is clear that regardless of any *philosophical* difference, *participants* do not take the proximal determinism descriptions to refer to the same thing as the distal determinism descriptions. Either one or both of the subscales must be failing to capture beliefs about determinism as commonly understood in the literature. I believe the best candidate for a more accurate description is that of distal determinism. In which case, the result provides evidence against the hypothesis that most people don't believe in either determinism or indeterminism. However, if the proximal description is more accurate, then the results support this hypothesis. So, the results are unclear.

I will now turn to a scale that makes significant improvements on previous scales.

The Free Will Inventory (FWI) developed by Nadelhoffer, Shepard, Nahmias, Sripada, and Ross (2014), after testing the effectiveness of a number of determinism items, settled on these five items:

> 1. Everything that has ever happened had to happen precisely as it did, given what happened before.

2. Every event that has ever occurred, including human decisions and actions, was completely determined by prior events.

3. People's choices and actions must happen precisely the way they do because of the laws of nature and the way things were in the distant past.

4. A supercomputer that could know everything about the way the universe is now could know everything about the way the universe will be in the future.

5. Given the way things were at the Big Bang, there is only one way for everything to happen in the universe after that. (pp. 34-35)

Participants were asked their level of agreement with each of the statements on a Likert scale from 1-(strong disagreement) to 7-(strong agreement). A good result for Vihvelin on such a scale would be a mean not significantly different from 4-(neither agree nor disagree). 330 participants completed the survey (p. 35). The means were 3.72, 3.44, 3.18, 2.97, and 3.10 respectively. [17] All have p-values less than 0.01 when compared to the neutral-point, which would mean rejecting the hypothesis that people neither agree nor disagree with the statements. Note also that *all* values are below 4. If some were above and some below then this could indicate confusion about the statements. If we treat all answers from 5-7 as "agreement", all from 1-3 as "disagreement", and 4 as "neither agree nor disagree" the results breakdown as follows:

1: 31% agree; 44% disagree; 25% neither agree nor disagree

2: 29% agree; 51% disagree; 20% neither agree nor disagree

3: 22% agree; 61% disagree; 17% neither agree nor disagree

4: 18% agree; 62% disagree; 20% neither agree nor disagree

5: 18% agree; 57% disagree; 25% neither agree nor disagree

This constitutes very strong evidence against Vihvelin's claim that there is no commonsense belief about determinism. However, we need to look for ways in which the experiment may have failed to properly gauge participants' actual beliefs.

Perhaps a similar critique to the one Nahmias (2006) and Vihvelin (2015) make of the Nichols and Knobe (2007) vignette could apply to Nadelhoffer et al.'s (2014) determinism item 1, "Everything that has ever happened had to happen precisely as it did, given what happened

---

[17] These are my calculations based on the full data set, which was not included in the published paper, but kindly shared with me by Thomas Nadelhoffer.

before" (p. 34). Determinism does not entail that everything "*had* to happen" without the qualifier "given what happened before". If participants either fail to notice or fail to understand the importance of this qualifier then they may interpret the statement as meaning simply "everything had to happen". If this was the case then they may disagree with the statement on the grounds that it implies no-one ever had the ability to do otherwise, even though determinism does not rule out the conditional analysis of "ability to do otherwise".

However, I do not think this is as plausible a problem here as it was for the Nichols and Knobe vignette. In that case the length of the text introduced several elements that could be interpreted in a number of ways, increasing the risk of the conditional being missed by the reader (see Vihvelin's (2015) bolding of these elements in Section 3.2.1). Nadelhoffer et al.'s item is a single short sentence, minimising the risk of the conditional being missed. On the other hand, shorter texts have the potential of *lacking enough* information to get the intended idea across. However, I think this is mitigated due to the use of multiple items for the same concept, as Nadelhoffer et al. do.

The language of other items could also be picked apart. Note that all the items attempt to convey the basic definition of determinism as it is commonly understood in the literature: that the state of the universe at any one time determines the state of the universe at all future times. I think the supercomputer item is the weakest in conveying this definition. It is plausible that participants will focus on the possibilities of technology (can a supercomputer *really* know *everything?*) rather than on the implications of such technology.

Whatever the limitations of these particular items, the onus is on the critic to formulate a hypothesis for why we get such results from such items. And if testing such a hypothesis proves these items to be inadequate, this can inform the development of superior items. As it stands, I believe these items constitute the clearest collection of determinism items in the literature.

The Free Will Inventory has been used in several studies since Nadelhoffer et al. (2014) published the original paper, including Nadelhoffer, Yin, and Graves (2020) and Nadelhoffer, Rose, Buckwalter, Nichols (2020). Both found substantial disagreement with the determinism items. Indeed, Nadelhoffer et al. (2020) presented the inventory to a large sample, some 1,385

participants, with mean results for each item even lower than they were in the Nadelhoffer et al. (2014) study.[18]

The Free Will Inventory has also been used in a large cross-cultural study (Wisniewski, Deutschlaender, Haynes, & Jong, 2019), comparing the beliefs of U.S.-based participants to those based in Singapore. This study had 900 participants in Singapore and 900 in the United States. The results for U.S. participants were consistent with the previous studies, finding that 30.77% believe in determinism. The results for the participants in Singapore was strikingly different, with a majority, 59.00%, believing in determinism (p. 8). This suggests that we should not assume that the majority of people are indeterminists just based on the data gathered so far. There is at least one population with a majority belief in determinism and it would be surprising if there were not at least some others. But this variation does not help Vihvelin's argument in any way. It seems that most people in Singapore have a belief about whether or not determinism is true, just as most people in the U.S. do. They just happen to disagree.

4.2 Beliefs on the Compatibility of Free Will and Determinism

I will now look at evidence regarding Vihvelin's second claim:

> 2. There is no commonsense belief about whether or not free will is compatible with determinism.

Again, we can take "commonsense belief" to mean "a belief held by most people". So, the hypothesis being tested here is:

> H2: Most people do not hold a belief about whether or not free will is compatible with determinism.

What we should expect if this hypothesis is true looks similar to our expectations for the previous claim. If participants are asked if they think a character in a deterministic scenario can act with free will then we would expect most people to respond that they "don't know". If given a binary choice, we would expect results not significantly different from 50/50.

Again, there may be reasons to query results. It's worth remembering that Vihvelin *expects* most people to have incompatibilist intuitions upon being exposed to philosophy. So,

---

[18] I calculated the means for the items to be: 1: 2.93, 2: 3.13, 3: 3.23, 4: 2.93, 5: 2.66. As the Nadelhoffer et al. (2020) FWI results were not pertinent to the main topic of the paper, the data is only available in the supplementary materials at https://osf.io/4z6rs

we need to be critical in trying to discern if the experiment is successful at getting at the beliefs people hold prior to the experiment.

Here I need to be careful around the use of the terms "incompatibilist" and "compatibilist". Up to this point I have been using the terms to refer to Vihvelin's definition, i.e., they both refer to people who claim free will is at least possible, and their claims regard the compatibility of free will and determinism, *not* moral responsibility and determinism. None of the empirical researchers I will look at used Vihvelin's definition of terms.

This doesn't make a lot of difference when it comes to the first part of the definition (that both sides claim free will is possible) because Vihvelin and the researchers are interested in what people believe about the compatibility question, not in their final positions. To see why this is the case, let's imagine a participant indicating a belief that free will is incompatible with determinism *and* a further belief that free will is impossible. While the researchers I look at would call such a person an incompatibilist (and perhaps a hard determinist), Vihvelin would not define them as an incompatibilist, but as an impossibilist. But when Vihvelin claims that there are no commonsense beliefs about the compatibility question, she is referring to their judgments about the question, not about their final claims regarding the possibility of free will. So, when a researcher says they have discovered most people are incompatibilists, they do so on the basis of participants' answers to the compatibility question. This means such a discovery would count as evidence against Vihvelin's claim – even if it might turn out that most of those people also think free will is impossible.

What *will* become an issue regards Vihvelin's strict concern with free will rather than moral responsibility. A number of the studies have tested beliefs about the compatibility of moral responsibility and determinism, taking this to indicate beliefs about the compatibility of free will and determinism. Vihvelin argues that these questions do not necessarily overlap. And there is research to back this up (Figdor & Phelan, 2015; Vierkant, Deutschländer, Sinnott-Armstrong, & Haynes, 2019). Unfortunately, though the number of studies that carefully distinguish between these questions is growing, we do not yet have the luxury of ignoring the studies that look only at moral responsibility.

4.2.1 Conflicting Early Results

This issue about the difference between the moral responsibility question and the free will question comes up in a pair of conflicting studies from the early experimental philosophy literature. Nahmias, Morris, Nadelhoffer, and Turner (2005) provide evidence for a majority

believing free will and determinism to be compatible, as well as blameworthiness and praiseworthiness being compatible with determinism. Nahmias et al. take the blameworthiness and praiseworthiness results to indicate compatibilist belief about moral responsibility. I will comment on this point below. The Nichols and Knobe (2007) study we looked at earlier found evidence for most people being *incompatibilists* with regard to moral responsibility when presented with an abstract scenario (no description of behaviour), but *compatibilists* when presented with a concrete scenario (description of an immoral act).

There were a number of differences between these studies that might account for the different results. Before I look at those though, I want to address the question of whether or not the Nichols and Knobe study can tell us anything about free will beliefs, given that it only asks participants about moral responsibility. Though there is evidence for judgments about free will being distinct from judgments about moral responsibility, the two do appear to track to some degree. Figdor and Phelan (2015) found free will judgments to be lower than moral responsibility judgments for deterministic scenarios, in almost all cases. But they were close enough such that in each case the average position of each judgment was the same. That is to say, where the mean judgment about moral responsibility indicated judging it incompatible with determinism, the mean judgment about free will also indicated an incompatibilist view (pp. 611-613).

Figdor and Phelan also found that judgments about blameworthiness and praiseworthiness did track moral responsibility judgments relatively closely (p. 611). This indicates that the Nahmias et al. (2005) results can also tell us something about how close moral responsibility judgments can be to free will judgments. Fitting the trend found by Figdor and Phelan, Nahmias et al. found free will judgments to be somewhat lower than judgments about blameworthiness and praiseworthiness, but close enough that both indicated a majority holding compatibilist beliefs. All of this suggests that the Nichols and Knobe (2007) study does give us *some* idea of what free will judgments might be, at least with regard to the scenarios they used.

However, given Nichols and Knobe find people make strikingly different judgments regarding concrete scenarios compared to abstract scenarios, it's plausible this effect is particular to moral responsibility judgments and not to free will judgments. Results from Feltz, Cokely, and Nadelhoffer (2009) suggest that this might actually be the case. Feltz et al. used the same descriptions of deterministic and indeterministic universes but asked about free will

instead of moral responsibility. When asked about whether someone acts of "his own free will" regarding an immoral act in a deterministic scenario, most participants gave incompatibilist responses (p. 11). However, the study did not include an abstract condition to contrast with the concrete cases. The moral acts were also different (rape and tax fraud, whereas Nichols and Knobe described a man murdering his family). Further, Nichols and Knobe (2007) express scepticism about whether asking if someone "acts of his own free will" gets at the intended target, as it's unclear if folk use of the term aligns with philosophical use (which is itself varied) (p. 682). This criticism seems to be borne out in results from Nahmias and Murray (2010) who tested the Nichols and Knobe vignettes, both abstract and concrete, with the term "free will", but not the phrase "acts of his own free will" (p. 201). In this study, the free will judgments in the abstract condition were higher than those for the concrete condition.[19] Bear and Knobe (2016) tested the same deterministic scenario with a variety of questions including ones about "willpower" and if someone could "change their mind" or "decide" differently (p. 2031). All three of these are aspects of the concept of free will Vihvelin (2013) has in mind (pp. 10, 15). In all three cases, the mean response was below the mid-point, meaning the tendency was to judge them all incompatible with determinism (Bear & Knobe, 2016, p. 2031).[20]

So, in at least some studies using the Nichols and Knobe vignettes, we get evidence for most people believing that *free will* is incompatible with determinism, even in concrete scenarios. But then what do we make of Nahmias et al. (2005) finding majority compatibilist views? One obvious explanation is that the studies used different descriptions of determinism. Nichols and Knobe (2007) described a universe in which "everything that happens is completely caused by whatever happened before it" (p. 669) (this is the vignette quoted in full in Chapter 2). Nahmias et al. (2005) had one scenario describing a supercomputer that can predict the future, and another scenario describing "a world where the beliefs and values of every person are caused completely by the combination of one's genes and one's environment" (p. 570). So, this is clearly an issue, but finding out *how* these different descriptions get interpreted by participants is not the domain of pure speculation. We can only get definitive results through testing hypothesised explanations. Many papers and studies have contributed to the debate on folk beliefs about the compatibility question (see Bear & Knobe, 2016; Cova, Bertoux, Bourgeois-Gironde, & Dubois, 2012; Cova & Kitano, 2014; Deery, Davis, & Carey,

---

[19] Unfortunately, there is a lack clarity as to just *how* different these judgments were as Nahmias and Murray (2010) only report a "composite score" calculated by averaging each participant's moral responsibility, free will, and blame judgments (p. 202).

[20] On a 7-point scale, the "willpower" item mean score was 2.86, "change mind" 2.93, and "decide" 3.71 (p. 2031).

2015; Feltz, Cokely, & Nadelhoffer, 2009; Mandelbaum, Ripley, Waytz, & Phelan, 2012; May, 2014; Murray & Nahmias, 2014; Nadelhoffer, Shepard, Nahmias, Sripada, & Ross, 2014; Nadelhoffer, Yin, & Graves, 2020; Nahmias, 2006; Nahmias, 2011; Nahmias, Coates, & Kvaran, 2007; Nahmias & Murray, 2010; Nichols, 2006; Rose, Buckwalter, & Nichols, 2017; Rose & Nichols, 2013; Sinnott-Armstrong, 2008; Turri, 2017; Turri, 2017; Woolfolk, Doris, & Darley, 2006; Shepherd, 2012). I will look at some of the most compelling now.

4.2.2 Affect, Bypassing, and Intrusive Indeterministic Metaphysics

Attempting to explain away compatibilist results, Nichols and Knobe (2007) put forward a hypothesis that people are naturally incompatibilists, but that *affect* causes a performance error. For example, descriptions of disturbing immoral acts, such as a man murdering his family, would cause participants to answer as compatibilists even though this did not reflect their true (incompatibilist) position. Nichols and Knobe (2007) found that this theory was supported by further studies.

The *affective performance error* model has some problems though. Nahmias et al. (2005) did ask about negatively valence acts without observing such a performance error (p. 567). And, as I went over in the previous section, the results for the Nichols and Knobe vignettes when participants are asked about *free will*, are inconsistent with regards to abstract vs. concrete conditions. So, it's not clear that affect, at least on its own, can explain away the Nahmias et al. (2005) results.

Nahmias (2006) proposed a competing theory to explain away *incompatibilist* intuitions. He argued that participants in the Nichols and Knobe study had been interpreting deterministic scenarios as implying that mental states have no effect on decisions – that mental states are *bypassed* (p. 217). Since it is widely agreed that determinism does *not* imply that mental states are not part of the causal chain leading to decision making, this would mean that participants were answering as incompatibilists for the wrong reason. Several studies have found evidence supporting the bypassing hypothesis (Nahmias, Coates, & Kvaran, 2007; Nahmias & Murray, 2010; Shepherd, 2012; Nahmias, Shepard, & Reuter, 2014).

However, Rose and Nichols (2013) present a very strong refutation of the bypassing hypothesis. The bypassing hypothesis relies on a particular causal process taking place, such that the description of determinism read by participants causes them to judge that mental states are bypassed, which in turn causes them to judge that free will and moral responsibility are not possible in the scenario. Rose and Nichols (2013) suggest an alternative direction of causation:

that the description of determinism causes the judgment that free will and moral responsibility are not possible, and it is *those* judgments that cause the bypassing judgments. Conducting a mediation analysis with structural equation models, Rose and Nichols (2013) were able to show that this alternative causal process was a strong fit for the data, whereas the bypassing model "did not fit the data at all" (p. 607). To drill home the significance of this finding, I should point out that the method Rose and Nichols used relies on an algorithm called the Greedy Equivalence Search (GES). GES has been proven by Chickering (2002) to return the *true causal model*, as long as certain conditions are met, such as sufficient data.[21] Further, this is not just evidence that the bypassing hypothesis fails to explain away incompatibilist judgments in earlier studies, it also provides very strong evidence that can explain away *compatibilist* judgments in (at least some of) the previous studies purporting to show bypassing.

I'll turn now to what I think is the most compelling hypothesis for explaining away compatibilist judgments – intrusive indeterministic metaphysics. Rose, Buckwalter, and Nichols (2017) suppose that if people believe human agency is indeterministic and if this belief is deeply ingrained, then people may respond to deterministic scenarios as if they were actually *indeterministic*. This would undercut the evidence for compatibilist belief. To test this view, they hypothesised that the representations participants have of some deterministic scenarios are "*filled in*" with their intuitive views of human agency (p. 484). They make a distinction between two ways scenarios can be filled in – "importing" and "intruding":

> *Importing* occurs when participants fill in the scenario in ways that are *consistent* with the scenario, but the filling-in systematically goes beyond the information provided in the scenario. Of course, when participants read vignettes, importing will be a common occurrence. It becomes theoretically interesting when the imported information undermines the interpretation of the results. *Intruding* occurs when the filling in leads to a *mis*representation of the scenario. (p. 484)

If participants fill in deterministic scenarios with their indeterministic intuitions, then this would be a case of intrusion.

Rose et al. (2017) used vignettes from Nahmias, Shepard, and Reuter (2014). Nahmias et al. (2014) had already established that participants respond to each scenario differently, attributing free will to someone who's voting decision is perfectly predicted by neuroscientists, but no free will when her decision is being manipulated by the neuroscientists. The purpose of Rose et al.'s (2017) experiments was to see if this can be explained by participants having

---

[21] For further discussion on GES, see Rose, Livengood, Sytsma, and Machery (2012).

deeply held indeterministic metaphysics intrude on their thinking about the scenarios. To test for this, in each experiment they asked a series of questions about each scenario, for example, whether the character could have changed her mind after the occurrence of the brain activity used for the neuro-prediction/manipulation and whether she could have changed her mind after becoming aware of who she was going to vote for. These questions represented different ways of testing how participants interpreted the deterministic nature of each scenario. Since both scenarios were designed to describe determinism and both used very similar phrasing, we would expect participants to answer consistently for each scenario if they were interpreting both as being deterministic. Instead, affirmative answers tended to be higher for the questions regarding the neuro-prediction scenario, than for the manipulation scenario (2017). This is evidence for people misrepresenting the neuro-prediction scenarios as indeterministic. That is, it is evidence for an intrusive indeterministic metaphysics.

This line of research was extended by Nadelhoffer, Rose, Buckwalter, and Nichols (2020). With a large sample size of 1,296 participants, they tested vignettes from Nahmias et al. (2005; 2006). One scenario involved a perfectly predicting supercomputer, the other a universe that is restarted again and again with the same results. For both scenarios Nadelhoffer et al. (2020) found evidence for intrusive indeterministic metaphysics. Participants were recruited through Amazon Mechanical Turk and were all located in the United States.

There is currently no research that undermines the evidence for intrusive indeterministic metaphysics. I take these results to be the best available evidence for explaining why there have been conflicting results on beliefs about the compatibility of free will and determinism. It undercuts the evidence for compatibilist belief and gives further weight to claims that people are naturally incompatibilists. This is evidence against the second hypothesis (that most people hold no belief about the compatibility of free will and determinism) and, as such, is evidence against Vihvelin's claim that there are no commonsense beliefs about the compatibility question.

It is also evidence against the first hypothesis, regarding beliefs about whether or not determinism is true, indicating Vihvelin is wrong in her claim that there are no commonsense beliefs about whether or not determinism is true. It strongly suggests that not only do people *have* a belief about whether determinism is true, but it is a *very deeply ingrained belief*. I think Nadelhoffer et al. are right when they say, "We believe that the main reason that intrusion effects are prevalent is that indeterminism about human agency is the default folk view" (p17).

4.2.3 Cross-Cultural Research

A major limitation of the preceding research is the vast bulk of it only studies the beliefs of people in the U.S. To get a sense of how beliefs about the compatibility of free will and determinism might vary across the globe I will now look at two cross-cultural studies.

The first, from Wisniewski, Deutschlaender, Haynes, and Jong (2019), has already been mentioned in Section 4.1.1. They found that belief in determinism varies between American and Singaporean participants (30.77% and 59.00% respectively), but that belief in free will was similar for both (82.33% and 85.44% respectively). Assuming that people are likely to be logically consistent in their beliefs, we can infer from these numbers that most Singaporean participants are compatibilists. And though one can believe in both indeterminism and compatibilism coherently, if we assume American participants see indeterminism as necessary for free will, then we can infer most American participants are incompatibilists. However, other results from the study suggest Americans may *not* see indeterminism as necessary, many agreeing with "compatibilist" items (p. 10).

However, I am sceptical that the "compatibilist" and "incompatibilist" items Wisniewski et al. used are quality measures of these beliefs. Take this item for example: "Free will is the ability to make a choice based on one's beliefs and desires such that, if one had different beliefs or desires, one's choice would have been different as well" (p. 7). Wisniewski et al. take it that agreement with this item indicates agreement with a conditional sense of the ability to do otherwise, and this in turn indicates compatibilist belief. But this isn't a necessary interpretation – one could very well believe in an *unconditional* ability to do otherwise and yet agree that choices probably *would* be different *if* beliefs and desires were different. The item fails to communicate that the *only* way choices could have been different in a deterministic world is if the past had been different. To get clear on this we would need to ask participants if it is possible to act *against* our beliefs and desires. Further, the evidence for intrusive indeterministic metaphysics suggests that these items may not be interpreted as deterministic. So, I think it's unclear what we should draw from this study regarding beliefs about the compatibility question.

The other cross-cultural study comes from Sarkissian, Chatterjee, De Brigard, Knobe, Nichols, and Sirker (2010), who found majority belief in the incompatibility of moral responsibility and determinism across four cultures. The obvious limitation here is that they asked about moral responsibility, *not* free will. However, they used the Nichols and Knobe

vignettes, which, as mentioned earlier, have also been found to elicit incompatibilist judgments about free will, albeit inconsistently. So, we may tentatively take the Sarkissian et al. study as evidence for majority incompatibilist views about free will and determinism in four cultures. But this is *very* tentative because it is entirely plausible that some cultures have concepts of "free will" and "moral responsibility" that diverge more significantly than they do in American culture. Clearly, more cross-cultural research is needed in this regard.

Overall though, the studies looked at in this section indicate that most people *do* have a belief about whether or not free will is compatible with determinism. It's not clear yet what proportion of people believe in incompatibilism compared to compatibilism, but the evidence on the whole tends to favour a majority belief in incompatibilism, at least among the American population. Regardless of the balance of these beliefs, that most people take a view *at all* indicates Vihvelin is wrong that there are no commonsense beliefs about the compatibility question.

4.3 Beliefs About the Ability to Do Otherwise

I will now look at the evidence regarding Vihvelin's third claim:

> 3. There is a commonsense belief that we have the ability to do otherwise, but they do not believe this is an unconditional ability.

Again, we can take "commonsense belief" to mean "a belief held by most people". So, the hypothesis we are testing here is:

> H3: Most people believe that they have the ability to do otherwise, but do not believe this is an unconditional ability.

Note that this leaves two positions that would support Vihvelin's claim: either most people believe in an ability to do otherwise and are neutral on whether or not it is a conditional ability, *or* they believe in a conditional ability to do otherwise. There are also two positions that would count as evidence against Vihvelin's claim: either most people believe in an *unconditional* ability to do otherwise, or they don't believe in an ability to do otherwise at all.

As before, I will look for ways that experiment design may fail to capture participants' prior beliefs.

A number of studies have been conducted on beliefs about the ability to do otherwise (Nahmias E. , Morris, Nadelhoffer, & Turner, 2004; Nahmias E. , Morris, Nadelhoffer, &

Turner, 2005; Nichols, 2004; Woolfolk, Doris, & Darley, 2006; Nichols, 2012; Deery, Bedke, & Nichols, 2013; Nahmias, Nadelhoffer, Schooler, & Vohs, 2015; Deery, Davis, & Carey, 2015; Nadelhoffer, Yin, & Graves, 2020).

The results across all studies support the claim that most people believe in an ability to do otherwise, though judgments can vary depending on context. For example, Nahmias et al. (2005) found that a majority judged someone to have the ability in regard to a negative moral behaviour but only a minority judged this was the case for positive and neutral behaviours (p. 569). But the bulk of the evidence confirms the first part of Vihvelin's claim; it does seem that a core commonsense belief about free will is that we have an ability to do otherwise. However, almost all studies found evidence that most people believe this is an *unconditional* ability. One that did not, Nahmias et al. (2004), was a small pilot study, so it's not clear much can be drawn from these results in light of subsequent larger studies. As it stands, only one study, Deery et al. (2015) seems to provide serious evidence for most people believing in a conditional ability to do otherwise.

Deery et al. (2015), found most participants agreed with both conditional *and* unconditional statements about an ability to do otherwise arguing that they showed a "clear preference" for the conditional ability (p. 791). Perhaps this characterisation is overly confident though. The mean agreement with the unconditional ability items, on a 7-point scale, was $M = 4.27$ ($SD = 1.06$) while mean agreement with the conditional ability items was $M = 5.06$ ($SD = 0.83$) (p. 791). This is a statistically significant difference, but I'm not sure it amounts to a *clear* preference. Nadelhoffer et al. (2020) also critique this point, noting: "If someone said that they're undecided about x and someone else says they somewhat agree with x, it's odd to conclude that the latter has a strong preference for x relative to the former" (p. 972).

So, while the bulk of the evidence is strongly against the third hypothesis (and hence Vihvelin's claim) since there were some conflicting results, it is worth trying to understand why that was the case. Once again, the evidence for an intrusive indeterministic metaphysics could help explain why a majority of participants chose "conditional" statements in the Nahmias et al. (2004) and Deery et al. (2015) studies despite the vast majority of studies finding evidence for the unconditional view. Perhaps most participants in these two studies actually believe in an unconditional ability, but misinterpreted the conditional statements as unconditional, due to an indeterministic metaphysics intruding on their understanding of the statements.

Finally, I will look at the findings in Nadelhoffer et al. (2020). They found that most participants do not attribute free will to agents in deterministic scenarios who only have a conditional ability to do otherwise, whereas most attribute free will to agents in *indeterministic* scenarios who have an *unconditional* ability to do otherwise (p. 975). This indicates a majority belief in both an unconditional ability to do otherwise and in the incompatibility of free will and determinism.

As noted in Chapter 3, evidence for most people believing in an unconditional ability to do otherwise undercuts all three of Vihvelin's claims. It shows she is only half-right about belief in the ability to do otherwise. Yes, people believe we have that ability, but they also believe it is an unconditional ability. It also suggests people have beliefs about both the truth of determinism and the compatibility of free will and determinism. In particular, it indicates belief in indeterminism and incompatibilism.

So, the available evidence indicates that all three hypotheses are false, suggesting that Vihvelin is wrong about three core claims regarding commonsense beliefs – all of which are important for her burden of proof argument. In the following chapter I will make clear how this evidence impacts her argument.

Chapter 5: Countering Vihvelin's Burden of Proof Argument

In Chapter 3 I formulated three claims Vihvelin makes about commonsense beliefs into testable hypotheses:

H_1: Most people do not hold a belief about whether or not determinism is true.

H_2: Most people do not hold a belief about whether or not free will is compatible with determinism.

H_3: Most people believe that we have the ability to do otherwise, but they do not believe this is an unconditional ability.

In Chapter 4 I looked at the empirical evidence in regard to these hypotheses. In this chapter I will show how this empirical data impacts Vihvelin's claims about commonsense beliefs and burden of proof.

5.1 How the Evidence Bears on Vihvelin's Claims About Commonsense Beliefs

Here I will look at how the evidence examined in Chapter 4 bears on the two main premises of Vihvelin's first assessment of argumentative burden. Remember that her argument goes: commonsense beliefs are neither determinist nor indeterminist, therefore commonsense beliefs are neither compatibilist nor incompatibilist. The first hypothesis looked at in Chapter 4 – regarding beliefs about whether not determinism is true – bears on the premise of this argument. The second and third hypotheses – on beliefs about the compatibility question and about the ability to do otherwise – both bear on the conclusion of the argument.

5.1.1 Beliefs About Determinism

Vihvelin claims that there is no commonsense belief about whether or not determinism is true. The evidence strongly indicates this is false. A number of high quality studies point to the majority of people in the U.S. believing in indeterminism (Nadelhoffer, Shepard, Nahmias, Sripada, & Ross, 2014; Nadelhoffer, Yin, & Graves, 2020; Nadelhoffer, Rose, Buckwalter, & Nichols, 2020; Wisniewski, Deutschlaender, Haynes, & Jong, 2019). Though it is not yet clear how prevalent indeterminist belief is outside the U.S., one cross-cultural study suggests this is also the dominant belief in three other cultures (Sarkissian, et al., 2010). Majority belief in

indeterminism may not be universal though, as Wisniewski et al.'s (2019) evidence suggests most people in Singapore believe in determinism. Much wider cross-cultural studies are needed in order to see how belief in determinism varies around the world. Though the majority view in the U.S. disagrees with the majority view in Singapore, the evidence is against the claim that the majority have no view on the matter. Now, it's *possible* that the U.S. and Singapore are outliers compared to the rest of the world, and that most people across the globe have no beliefs about determinism, but that seems *very* implausible.

Further, the evidence for intrusive indeterministic metaphysics (Rose, Buckwalter, & Nichols, 2017; Nadelhoffer, Rose, Buckwalter, & Nichols, 2020) suggests that not only do most people have a view, but that view, at least for Americans, is *very* deeply ingrained.[22] This is noteworthy because it gets at an important part of Vihvelin's conception of common sense. Vihvelin's view is probably most closely aligned with that of Moore. This is made most clearly in Vihvelin's (2011) paper, which was heavily drawn on for Chapter 1 of her (2013) *Causes, Laws, and Free Will*. Vihvelin (2011) describes the belief in the ability to do otherwise as a "Moorean fact" (pp. 324-325). These are facts that are "obviously and undeniably true" (p. 325), so much so that "we typically don't bother saying [them]" (p. 336).[23] It seems quite plausible that those who's indeterministic metaphysics is so deeply ingrained that it intrudes on judgments about deterministic scenarios also take it that indeterminism is obviously and undeniably true. Of course, this needs to be studied to be sure. But I think it's the best evidence we've got so far that many people, or at least among Americans, might take indeterminism to be something like a Moorean fact.

So, though the evidence is incomplete, I think we can feel very confident that Vihvelin is wrong about there being *no* commonsense belief on whether or not determinism is true.

5.1.2 Beliefs About the Compatibility of Free Will and Determinism and the Ability to Do Otherwise

Vihvelin claims there are no commonsense beliefs about whether or not free will and determinism are compatible. This was more difficult to assess than the first claim because there

---

[22] There have not yet been cross-cultural studies on intrusive indeterministic metaphysics. Perhaps such studies would find that Singaporeans are prone to intrusive indeterministic thinking and that more are indeterminists than the Wisniewski et al. (2019) study suggests. Or it may uncover that they are less prone to this kind of intrusion than Americans, providing even stronger evidence that most Singaporeans are determinists.

[23] It's worth noting that despite heavy use of parts of Vihvelin (2011) in Vihvelin (2013), "Moorean fact" is a prominent part of the former, while being completely avoided in the latter. This may indicate a change in views, but if that's true, it's unclear to what extent. Regardless, Vihvelin (2013) continues to use Moorean ways of describing core commonsense beliefs such as that they are "obviously and undeniably true" (p. 15).

were a number of conflicting results and competing explanations for this conflict. However, the best evidence at this stage does *not* lend support for Vihvelin's claim. Rose and Nichols (2013) demonstrate that the belief that a scenario is deterministic *causes* people to judge free will would not be possible. This study also brings into doubt many results for compatibilist judgments, suggesting the dominant belief, in the U.S. at least, is incompatibilist. The intrusive indeterministic metaphysics results also point toward resolving the conflicting results in favour of the argument that most people (or at least Americans) are incompatibilists. We also see tentative results for incompatibilist belief in three other cultures with the Sarkissian et al. (2010) study that looked at beliefs about the compatibility of moral responsibility and determinism.

Though I was unable to look at evidence regarding whether or not people have *considered* determinism, it's worth speculating about Vihvelin's (2013) alternative claim: that they *have* considered determinism, but have "never taken [it] seriously" (p. 29). The intrusive indeterministic metaphysics studies showed that many participants fail to comply with the conditions of deterministic scenarios in thought experiments. This could be taken as evidence that people don't take determinism seriously. But Vihvelin needs the lack of seriousness to result in people not coming to a view on the implications determinism might have for free will, and the evidence is that they *do* take determinism seriously enough to come to a view.

The abundance of evidence for most people believing in an unconditional ability to do otherwise (Nahmias E. , Morris, Nadelhoffer, & Turner, 2005; Nichols, 2004; Woolfolk, Doris, & Darley, 2006; Nichols, 2012; Deery, Bedke, & Nichols, 2013; Nahmias, Nadelhoffer, Schooler, & Vohs, 2015; Nadelhoffer, Yin, & Graves, 2020) also tells us something about commonsense beliefs regarding the compatibility question. Assuming that people are consistent in their views, believing the ability to do otherwise is only possible in an indeterministic world indicates a belief that free will and determinism are incompatible.

5.2 Reassessing Burden of Proof

Let's run through Vihvelin's (2013) process for applying argumentative burdens again. Vihvelin's first assessment of burden involves appealing to commonsense belief. A burden is applied to those who either disagree with a commonsense belief or those who make a claim for which there is no commonsense belief. In the debate between the compatibilist and the incompatibilist, Vihvelin claims there are no commonsense beliefs on the matter, so both have

a burden applied, neither having a higher burden than the other.[24] Vihvelin's second assessment of burden involves looking at the modal claims made by each side of the debate. Since the incompatibilist makes a more restrictive modal claim than the compatibilist, a further burden is applied to the incompatibilist, resulting in a higher burden for the incompatibilist overall. Taking Vihvelin's assumptions about how to assess burden as given, in this section I will reassess the burdens in light of the evidence.

5.2.1 Reassessing the Burden Based on Commonsense Beliefs

The evidence for beliefs about whether or not determinism is true means we should reject the premise of Vihvelin's argument for there being no commonsense beliefs about the compatibility question. She does not have an empirical basis from which to draw this inference. But that, taken by itself, doesn't necessarily mean there is a commonsense belief on the question. Vihvelin (2013) notes that people can believe in indeterminism without believing it is necessary for free will (p. 128). However, since there is also ample evidence for beliefs about the compatibility question, we can also reject this part of the argument.

The evidence indicates that most Americans are incompatibilists. If we assess burden based on those beliefs, then the burden is clearly on the compatibilist. However, we also have evidence for most Singaporeans being determinists. We don't have direct evidence about their beliefs on compatibility, but we do have evidence that most believe in free will.[25] So, assuming consistency in their beliefs, they are effectively compatibilists. In this case, the burden lies on the incompatibilist. But this brings up an important question: does it make sense to have *different* assessments of burden of proof for different cultures?

I think it does, because these results suggest there is not a single concept at play. Call the concept held by most Americans FREE WILL-A and that held by most Singaporeans FREE WILL-S. The evidence suggests that FREE WILL-A is incompatible with determinism, therefore the burden of proof is on the compatibilist *about FREE WILL-A* to tell us why this concept is in fact compatible with determinism. The evidence also tells us that FREE WILL-S is compatible with determinism (assuming consistency), therefore the burden of proof is on the incompatibilist about FREE WILL-S to tell us why this is false about *this* concept.

---

[24] As noted in Chapter 2, Vihvelin initially puts this as "neither side has the burden of proof", but later makes it clear that she means *both* sides have a burden, just that neither has the *higher* burden on the basis of commonsense belief alone.

[25] Or at least we don't have *good* direct evidence (return to Section 4.2.3 for discussion).

With this is in mind, it's worth noting that most studies on compatibility judgments found a significant minority disagree with the majority. It's unclear if the compatibilist minority in the American population have the *same* concept as the compatibilist majority in the Singaporean population (likewise for each respective incompatibilist portions). But whether this is the case or not, in both populations there seem to be at least two different concepts. Again, where the burden lies needs to be made specific to *which* concept is under debate.

I think we can confidently assume that the commonsense beliefs Vihvelin have in mind are those that are central to the American concept of free will. This seems clear given that she often refers to the free will beliefs of her students, who she presumably meets where she teaches at the University of Southern California, and given that she is communicating largely with the Anglo-American philosophical community. The evidence strongly indicates that this concept is what I've called FREE WILL-A, an *incompatibilist* concept. This indicates Vihvelin is wrong about where the burden lies when it comes to the commonsense beliefs regarding the concept she is concerned with. The evidence tells us that the *compatibilist* about FREE WILL-A has the burden of proof, and the incompatibilist *does not* (at least as far as commonsense beliefs are concerned).

5.2.2 Reassessing the Burden Based on Possibility Claims

I will now look at how Vihvelin's second source of burden comes into play. Her argument here is that those who claim something is impossible have a higher argumentative burden than those who claim it is possible. Regardless of what the commonsense beliefs are, the incompatibilist makes a claim of impossibility whereas the compatibilist makes a claim of possibility. So, does that mean that we simply place a burden on the incompatibilist, regardless of what the commonsense beliefs are? I think not.

Vihvelin describes burdens being placed and discharged at various stages in a dialectic between each side of the debate. If her claims about commonsense beliefs are correct, then the first stage of the dialectic simply fails to establish burden. It's the second stage of the dialectic at which the appeal is made to possibility claims. Since both sides need to make an argument, the assessment at the second stage is about which argument has the higher burden, the conclusion in this case being the incompatibilist.

Now let's see what happens in light of the evidence. Working with FREE WILL-A again, the compatibilist has argumentative burden at the first stage of the dialectic, while the incompatibilist has no burden. So, the second stage of the dialectic is now just the compatibilist

making their case. Once they discharge their burden through making their case, the burden is on the incompatibilist to show why the compatibilist is wrong. And so the dialectic goes on. But can we appeal to possibility claims to suggest that the argument required by the compatibilist at this second stage of the debate is any easier than the one required by the incompatibilist at the following stage? Perhaps, but note that the burden placed on the compatibilist at the first stage is of some significance. The compatibilist is now arguing against empirical evidence about what the concept FREE WILL-A is. To discharge their burden, they need to show that the evidence has not properly captured the concept. In most cases this will require empirical evidence for whatever hypothesis is being proposed. At this point the fact that the compatibilist claims something is possible seems not as important to the larger task they are burdened with. Then, once they've discharged their burden, it is up to the incompatibilist to demonstrate what the compatibilist got wrong. Again, this is likely to require empirical support. The fact that the incompatibilist is making an impossibility claim seems to be a minor matter in comparison to the task of demonstrating, empirically, what the concept actually is.

Chapter 6: Wider Implications

Here I will look briefly at the implications these results have for Vihvelin's broader argument as well as arguments about free will and burden of proof more generally.

6.1 Implications for Vihvelin's Commonsense Compatibilism

So far, I have outlined how the empirical evidence gives us strong reasons to reject Vihvelin's (2013) argument that the higher burden of proof is on the incompatibilist. In this section I will look at the implications this has for other arguments she makes in *Causes, Laws, and Free Will*.

6.1.1 The Failure of Commonsense Compatibilism

Vihvelin calls the kind of compatibilism she defends "commonsense compatibilism", since it is meant to align with commonsense beliefs about free will (p. 33). In her penultimate chapter, Vihvelin goes through the steps she has taken up to that point to complete her defence of commonsense compatibilism. These steps include her claims about commonsense belief and burden of proof, as well as counters to various fatalist arguments, Frankfurt cases and incompatibilist arguments. Here is how she puts the contribution of her burden of proof claims to her overall argument:

> I argued that in the debate between the impossibilist and the possibilist, the burden of proof is on the impossibilist because the impossibilist is arguing against the claims of common sense. I argued that this burden has not been met, and that there are reasons for believing it cannot be met. I also argued that, so far as the debate about the Determinism question is concerned, the compatibilist discharges her burden of proof by carefully describing a case in which someone has and exercises free will in a way that shows it to be at least prima facie possible and also compatible with determinism. I have described a number of such cases over the course of this book. The dialectical situation, then, is this. It is the incompatibilist's burden to produce a sound argument for the claim that there is some hidden contradiction or other impossibility in any such story. More generally, the incompatibilist needs an argument that no deterministic world is a Free Will world. But a close scrutiny of the best arguments for incompatibilism has revealed no such argument. The case for compatibilism is now closed. In the absence of any *new* arguments for incompatibilism, we are under no obligation to suspend belief or to retain an open mind; we are entitled to believe, and should believe, that free will is not only possible and compossible with determinism but that we *actually* have the free will that common sense says we have. We are entitled to answer "yes" to all three of my questions about free will—the Possibility question, the Determinism question, and the Existential question. (pp. 167-168)

This highlights the importance of Vihvelin's commonsense belief claims and the burden of proof claims she derives from them. If she's wrong about those then she cannot complete her defence of compatibilism. I have shown she *is* wrong, that the evidence shows commonsense beliefs about free will reveal at least two concepts, which I've dubbed FREE WILL-A and FREE WILL-S. Since FREE WILL-A is the dominant concept of free will in the U.S., it is also very likely that this is the concept Vihvelin is attempting to understand. And the evidence strongly indicates that FREE WILL-A is decidedly *incompatibilist*. So, it's not just that she fails to complete her defence of compatibilism, she fails to even get it off the *ground*. Her arguments against various incompatibilist accounts are irrelevant, because there is strong empirical evidence that the concept being debated, according to the ordinary people who hold and employ it in their daily lives, i.e., *according to common sense*, simply *is incompatible with determinism*. As long as we're talking about the dominant concept of free will in the U.S., there is no hope for a "commonsense compatibilism".

6.1.2 The Bundle of Dispositions View

Vihvelin follows her declaration of victory with a complex and interesting account of "commonsense compatibilism" that she calls the "Bundle of Dispositions view" (p. 169). Without going into too much detail, this is essentially an account of how a conditional analysis of the ability to do otherwise may be successful if agents have a bundle of dispositions of abilities that allow for certain kinds of counterfactuals to obtain in deterministic possible worlds.[26]

Since the evidence shows there is no way *any* account of commonsense compatibilism can work in regard to the concept at hand, FREE WILL-A, the Bundle of Dispositions view cannot be regarded as a commonsense account. This doesn't mean it needs to be thrown out though. As long as it is acknowledged that this account is *not* an attempt to show that the commonsense beliefs about FREE WILL-A are compatible with determinism, the account could be put to use in other ways. For example, it could be a revisionist account. Such a revisionist argument might go as follows: despite FREE WILL-A including an unconditional analysis of the ability to do otherwise, we would be better off if the concept was revised to have a conditional analysis. Or the Bundle of Dispositions view could be put to use as a potential account of FREE WILL-S or of the apparently compatibilist concept held by a

---

[26] Vihvelin calls her analysis LCA-PROP, after the two analyses she says it is a combination of: Lewis's Reformed Conditional Analysis and Manley and Wasserman's PROP analysis. (p. 183).

minority of Americans. And these two routes could be combined, i.e., one argument in favour of revisionism could be that there are already people who hold a concept of free will that is compatible with determinism, so it's feasible for *more* to do so.

The revisionist route should be taken with caution though. In light of the intrusive indeterministic metaphysics results, there is a high risk that those who hold FREE WILL-A will misunderstand Vihvelin's argument for a conditional analysis of the ability to do otherwise. Their indeterministic metaphysics could intrude on their reading of her descriptions of deterministic possible worlds. So, though a revisionist Vihvelin may *seem* to convince many readers to revise their concept, that may be on the basis of false assumptions about what such revision entails.

6.2 Implications for claims About Burden of Proof in Free Will Arguments Generally

6.2.1 The Burden on All Sides of the Free Will Debate

The evidence obviously also has implications for any argument that appeals to commonsense beliefs (or popular intuition, opinion and so on). Anyone appealing to the common view for establishing the burden of proof in the debate between the compatibilist and the incompatibilist must get clear that they are making an empirical claim and take the available evidence seriously. It is incumbent on them to cite the evidence for or against their arguments. The compatibilist needs to make clear which concept they have in mind and, if it is FREE WILL-A, they need to make clear why they think the empirical evidence for the concept being incompatibilist does not stack up. For a serious rebuttal, they should make testable hypotheses, and then test them – or at least convince someone else to test them. This doesn't mean incompatibilists should rest easy. They too must make clear that they are making an empirical claim, make clear what concept that claim is about and point to the evidence that backs it up.

Indeed, it would be negligent for philosophers of free will to make *any* kind of appeal to folk beliefs without checking the empirical literature to see if there is evidence as to what those beliefs happen to be. Even in cases where the evidence is still rudimentary, any philosopher who takes a scientific approach seriously should acknowledge the evidence that is there and point to the need for further research (again, perhaps even attempting to *do* some of that research, or at least trying to convince experimenters to undertake it).

This may come across as a bit heavy handed. This argument is obviously not aimed at those already taking the available empirical data seriously. Unfortunately, many who appeal to

common beliefs are *not* citing the evidence. Here are two recent examples. Jaster (2020), makes a series of claims about how ordinary language use of "free" is context dependent, yet cites no empirical research to that effect. This is not to say Jaster is *wrong*, but we need better evidence than speculation to move the debate forward. Jeppsson (2020), presents an argument against Pereboom's (2001) Four-Case Manipulation Argument, which includes questioning whether it's possible to take into account both Plum's agency and the manipulation (as Pereboom insists we should) when considering the implications of the thought experiment. There is actually research on this very thought experiment (Sripada, 2012), some of the results of which could be very useful for Jeppsson's argument.

21st Century Philosophy needs to get into the habit of dealing with claims about "ordinary language", "commonsense beliefs", "pre-theoretic intuitions" and the like, as *empirical claims*, and justify such claims on an empirical basis, insofar as they *can* be justified.

6.2.2 Revisionism

The problem intrusive indeterministic metaphysics brings to a revisionist version of Vihvelin's Bundle of Dispositions view is a potential problem for all revisionist projects. If adherents import indeterminism into compatibilist accounts, they may come away with a great deal of confidence in the coherence of compatibilism while believing mistaken things about it (e.g., that it supports an unconditional ability to do otherwise). Persuading people with intrusive indeterministic metaphysics to become compatibilists *and* understand what that actually amounts to is likely a very difficult task. Not only must you argue against the common view, but you must argue against a deeply entrenched view. And even if someone properly understands the difference between a concept of free will that allows only for a conditional ability to do otherwise and one that requires an unconditional ability, it is likely a very hard task for them to maintain that view in their everyday life.

There is reason for revisionists to be optimistic though. Given that there does appear to be compatibilist concepts of free will in use by significant numbers of people, such as FREE WILL-S, studying how it is that such people make sense of ideas like the ability to do otherwise might be informative for those wishing to revise FREE WILL-A. Over the long term, the empirical data can also tell us how concepts might evolve over time. That is to say, once experimental materials and design is of a sufficient standard, we may interpret variations in results to reflect *changes* in concepts. A revisionist should be very interested in finding out what kinds of things drive individuals and populations to arrive at a changed concept.

6.3 Follow up Research

There are many avenues of research yet to be taken up on questions about folk free will beliefs. Here I will briefly propose some that come to mind in light of the results we've looked at.

Clearly more cross-cultural research is needed. This goes not only for cultures that we might *expect* to be different from the U.S. (non-English speaking, Eastern, and so on), but for cultures we might expect to be similar, such as other English-speaking, Western countries. We very well might find some surprising differences.

Extending the intrusive indeterministic metaphysics research would be valuable. For example, looking at how Singaporeans respond to such research could tell us more about why so many report determinist beliefs.

It could also be informative to research what people find "obviously and undeniably true" about free will – the commonsense beliefs Vihvelin (2011) calls "Moorean facts". Further, it would be interesting to see if such beliefs have intrusive effects.

Perhaps Vihvelin is right that people do not *consider* determinism (prior to being presented with studies). If that can be shown, even about vague concepts of determinism, it could have implications for how we read the data. Nichols (2015) takes a number of the empirical results to indicate folk commitment to incompatibilist belief, *despite* them probably never having heard of causal determinism. How might his "flatfooted" response look if most people have never even considered a non-technical conception of determinism?

Chapter 7: Conclusion

Vihvelin's argument for the incompatibilist having the higher argumentative burden fails. I have demonstrated this by attacking her argument at every step. I have shown that this is an empirical question, and that scepticism of philosophical intuitions does not warrant rejection of empirical results. I have carefully examined the available data, demonstrating that Vihvelin is wrong in all her claims about commonsense beliefs that are crucial to her burden of proof argument. She is wrong about a lack of commonsense beliefs regarding whether or not determinism is true. She is wrong about a lack of commonsense beliefs regarding the compatibility of free will and determinism. She is wrong that people do not believe in an unconditional ability to do otherwise. I have shown that these results lead to the collapse of Vihvelin's burden of proof argument. I have looked at how these results impact Vihvelin's argument for commonsense compatibilism and her Bundle of Dispositions view, showing that she is forced to change direction, perhaps to a revisionist approach. Finally, I have drawn out some of the implications for the wider free will debate, urging more philosophers to take an empirical approach to these questions seriously.

I will end by pressing again on this final point, if I can be permitted to belabour the point. Vihvelin is far from alone in claiming to know what the beliefs of ordinary people are without any recourse to evidence. Making armchair claims about ordinary beliefs is a tradition as old as philosophy itself. But philosophers have been well aware of the *inadequacy* of armchair speculation about common beliefs for as long as such speculations have been made. Take this passage from Aristotle for example:

> reasoning is "contentious" if it starts from opinions that seem to be generally accepted, but are not really such, or again if it merely seems to reason from opinions that are or seem to be generally accepted. For not every opinion that seems to be generally accepted actually is generally accepted. (Topics, 100b24)

We would hope, given the invention of statistical methods and the development of experimental psychology, that contemporary philosophers would move on from making *guesses* about ordinary beliefs and actually *find out* what they are. Thankfully, the experimental philosophy movement has shown that more and more philosophers are willing to do just that.

The debate about the compatibility of free will and determinism is a prime example of the kind of debate that can be enlightened by experimental philosophy and other empirical research focussed on philosophical questions. Indeed, insofar as the compatibility debate is about the popular concept(s) of free will, the ongoing dialectic from nearly 20 years of experimental philosophy of free will *is the debate*.

So, rather than "avoid experimental philosophy" as Vihvelin (2015, p. 398) advises, if we are ever to decisively answer the compatibility question, *we cannot do without it*.

Bibliography

Aristotle. (1955). Topics. In W. D. Ross (Ed.), *The Works of Aristotle* (W. A. Pickard-Cambridge, Trans., Vol. 1, p. 100b24). London: Oxford University Press.

Bear, A., & Knobe, J. (2016). What Do People Find Incompatible With Causal Determinism? *Cognitive Science, 40*, 2025-2049. doi:10.1111/cogs.12314

Chickering, D. M. (2002). Optimal Structure Identification With Greedy Search. *Journal of Machine Learning Research, 3*, 507-554.

Cova, F., & Kitano, Y. (2014). Experimental Philosophy and the Compatibility of Free Will and Determinism: A Survey. *Annals of the Japan Association for Philosophy of Science, 22*, 17-37. doi:10.4288/jafpos.22.0_17

Cova, F., Bertoux, M., Bourgeois-Gironde, S., & Dubois, B. (2012). Judgments about moral responsibility and determinism in patients with behavioural variant of frontotemporal dementia: Still compatibilists. *Consciousness and Cognition, 21*(2), 851-864. doi:10.1016/j.concog.2012.02.004

Deery, O., Bedke, M., & Nichols, S. (2013). Phenomenal Abilities: Incompatibilism and the Experience of Agency. In D. Shoemaker (Ed.), *Oxford Studies in Agency and Responsibility* (Vol. 1, pp. 126-150). Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199694853.001.0001

Deery, O., Davis, T., & Carey, J. (2015). The Free-Will Intuitions Scale and the question of natural compatibilism. *Philosophical Psychology, 28*(6), 776-801. doi:10.1080/09515089.2014.893868

Ekstrom, L. W. (2005). Libertarianism and Frankfurt-style Cases. In R. Kane (Ed.), *The Oxford Handbook of Free Will* (1 ed., pp. 309-322). Oxford: Oxford University Press.

Feltz, A., & Millan, M. (2015). An error theory for compatibilist intuitions. *Philosophical Psychology, 28*(4), 529-555. doi:10.1080/09515089.2013.865513

Feltz, A., Cokely, E. T., & Nadelhoffer, T. (2009). Natural Compatibilism versus Natural Incompatibilism: Back to the Drawing Board. *Mind & Language, 24*, 1-23. doi:doi.org/10.1111/j.1468-0017.2008.01351.x

Figdor, C., & Phelan, M. (2015). Is Free Will Necessary for Moral Responsibility?: A Case for Rethinking Their Relationship and the Design of Experimental Studies in Moral Psychology. *Mind & Language, 30*, 603-627. doi:10.1111/mila.12092

Inwagen, P. v. (1983). *An Essay on Free Will.* Oxford: Clarendon Press.

Jaster, R. (2020). Contextualizing Free Will. *Zeitschrift für Philosophische Forschung, 74*(2), 187-204. doi:10.3196/004433020829410460

Jeppsson, S. (2020). The agential perspective: a hard-line reply to the four-case manipulation argument. *Philosophical Studies, 177*(7), 1935-1951. doi:10.1007/s11098-019-01292-2

Kane, R. (1999). Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism. *The Journal of Philosophy, 96*(5), 217-240. doi:10.2307/2564666

Lycan, W. G. (2003). Free Will and the Burden of Proof. *Royal Institute of Philosophy Supplement, 53*, 107-122. doi:10.1017/S1358246100008298

Mandelbaum, E., Ripley, D., Waytz, A., & Phelan, M. (2012). Explaining the Abstract/Concrete Paradoxes in Moral Psychology: The NBAR Hypothesis. *Review of Philosophy and Psychology, 3*(3), 351-368. doi:dx.doi.org/10.1007/s13164-012-0106-3

May, J. (2014). On the very concept of free will. *Synthese, 191*(12), 2849-2866. doi:10.1007/s11229-014-0426-1

Mickelson, K. (2015). A critique of Vihvelin's Three-fold Classification. *Canadian Journal of Philosophy, 45*(1), 85-99. doi:10.1080/00455091.2015.1009321

Morris, S. G. (2015). Commentary on "The Free-Will Intuitions Scale and the Question of Natural Compatibilism". *Philosophical Psychology, 28*(6), 802-807. doi:10.1080/09515089.2014.926440

Murray, D., & Nahmias, E. (2014). Explaining Away Incompatibilist Intuitions. *Philosophy and Phenomenological Research, 88*(2), 434-467. doi:10.1111/j.1933-1592.2012.00609.x

Nadelhoffer, T., Rose, D., Buckwalter, W., & Nichols, S. (2020). Natural Compatibilism, Indeterminism, and Intrusive Metaphysics. *Cognitive Science, 44*, e12873. doi:doi.org/10.1111/cogs.12873

Nadelhoffer, T., Shepard, J., Nahmias, E., Sripada, C., & Ross, L. T. (2014). The free will inventory: Measuring beliefs about agency and responsibility. *Consciousness and Cognition, 25*, 27-41. doi:10.1016/j.concog.2014.01.006

Nadelhoffer, T., Yin, S., & Graves, R. (2020). Folk intuitions and the conditional ability to do otherwise. *Philosophical Psychology, 33*(7), 968-996. doi:10.1080/09515089.2020.1817884

Nahmias, E. (2006). Folk Fears about Freedom and Responsibility: Determinism vs. Reductionism. *Journal of Cognition and Culture, 6*(1-2), 215-237. doi:10.1163/156853706776931295

Nahmias, E. (2011). Intuitions about Free Will, Determinism, and Bypassing. In R. Kane (Ed.), *The Oxford Handbook of Free Will* (2 ed., pp. 555-576). New York: Oxford University Press. doi:10.1093/oxfordhb/9780195399691.001.0001

Nahmias, E., & Murray, D. (2010). Experimental Philosophy on Free Will: An Error Theory for Incompatibilist Intuitions. In J. Aguilar, A. Buckareff, & K. Frankish (Eds.), *New Waves in Philosophy of Action* (pp. 189-215). Palgrave-Macmillan.

Nahmias, E., Coates, D. J., & Kvaran, T. (2007). Free Will, Moral Responsibility, and Mechanism: Experiments on Folk Intuitions. *Midwest studies in philosophy, 31*(1), 214-242. doi:10.1111/j.1475-4975.2007.00158.x

Nahmias, E., Morris, S. G., Nadelhoffer, T., & Turner, J. (2004). The Phenomenology of Free Will. *Journal of Consciousness Studies, 11*(7-8), 162-179.

Nahmias, E., Morris, S. G., Nadelhoffer, T., & Turner, J. (2006). Is Incompatibilism Intuitive? *Philosophy and Phenomenological Research, 73*(1), 28-53. doi:10.1111/j.1933-1592.2006.tb00603.x

Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2005). Surveying Freedom: Folk Intuitions about free will and moral responsibility. *Philosophical Psychology, 18*(5), 561-584. doi:10.1080/09515080500264180

Nahmias, E., Nadelhoffer, T., Schooler, J., & Vohs, K. D. (2015). Measuring and Manipulating Beliefs and Behaviors Associated with Free Will. In A. R. Mele (Ed.), *Surrounding Free Will: Philosophy, Psychology, Neuroscience* (pp. 72-94). New York: Oxford University Press. doi:10.1093/acprof:oso/9780199333950.003.0005

Nahmias, E., Shepard, J., & Reuter, S. (2014). It's OK if 'my brain made me do it': People's intuitions about free will and neuroscientific prediction. *Cognition, 133*(2), 502-516. doi:10.1016/j.cognition.2014.07.009.

Nettler, G. (1959). Cruelty, Dignity, and Determinism. *American Sociological Review, 24*(3), 375-384. doi:10.2307/2089386

Nichols, S. (2004). The Folk Psychology of Free Will: Fits and Starts. *Mind & Language, 19*, 473-502. doi:10.1111/j.0268-1064.2004.00269.x

Nichols, S. (2006). Folk Intuitions on Free Will. *Journal of Cognition and Culture, 6*(1-2), 57-86. doi:10.1163/156853706776931385

Nichols, S. (2012). The Indeterminist Intuition: Source and Status. *The Monist, 95*(2), 290-307.

Nichols, S. (2015). *Bound: Essays on free will and responsibility.* Oxford: Oxford University Press.

Nichols, S., & Knobe, J. (2007). Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions. *Noûs, 41*(4), 663-685. doi:10.1111/j.1468-0068.2007.00666.x

Paulhus, D. L., & Carey, J. M. (2011). The FAD–Plus: Measuring Lay Beliefs Regarding Free Will and Related Constructs. *Journal of Personality Assessment, 93*(1), 96-104. doi:10.1080/00223891.2010.528483

Paulhus, D. L., & Margesson, A. (1994). Free will and scientific determinism (FAD-4) scale. *Unpublished instrument, University of British Columbia, Vancouver, BC, Canada.*

Pereboom, D. (1995). Determinism al Dente. *Noûs, 29*(1), 21-45. doi:10.1080/00455091.2015.1009321

Pereboom, D. (2001). *Living Without Free Will.* Cambridge: Cambridge University Press.

Pink, T. (2004). *Free Will: A Very Short Introduction.* Oxford: Oxford University Press.

Reid, T. (1823). *An Inquiry into the Human Mind on the Principles of Common Sense.* Cupar: Tullis.

Rose, D., & Nichols, S. (2013). The Lesson of Bypassing. *Review of Philosophy and Psychology, 4*(4), 599-619. doi:dx.doi.org/10.1007/s13164-013-0154-3

Rose, D., Buckwalter, W., & Nichols, S. (2017). Neuroscientific Prediction and the Intrusion of Intuitive Metaphysics. *Cognitive Science, 41*, 482-502. doi:10.1111/cogs.12310

Rose, D., Livengood, J., Sytsma, J., & Machery, E. (2012). Deep trouble for the deep self. *Philosophical Psychology, 25*(5), 629-646. doi:10.1080/09515089.2011.622438

Sarkissian, H., Chatterjee, A., Brigard, F. D., Knobe, J., Nichols, S., & Sirker, S. (2010). Is Belief in Free Will a Cultural Universal? *MInd & Language, 25*, 346-358. doi:10.1111/j.1468-0017.2010.01393.x

Shepherd, J. (2012). Free Will and Consciousness: Experimental Studies. *Consciousness and Cognition, 21*(2), 915-927. doi:10.1016/j.concog.2012.03.004

Sinnott-Armstrong, W. (2008). Abstract + Concrete = Paradox. In J. Knobe, & S. Nichols (Eds.), *Experimental Philosophy* (pp. 209-230). New York: Oxford University Press.

Sripada, C. S. (2012). What Makes a Manipulated Agent Unfree? *Philosophy and Phenomenological Research, 85*, 563-593. doi:10.1111/j.1933-1592.2011.00527.x

Strawson, G. (2010). *Freedom and belief* (2 ed.). Oxford: Oxford University Press.

Stroessner, S. J., & Green, C. W. (1990). Effects of Belief in Free Will or Determinism on Attitudes Toward Punishment and Locus of Control. *The Journal of Social Psychology, 130*(6), 789-799. doi:10.1080/00224545.1990.9924631

Talbot, B. (2009). *How to use intuitions in philosophy.* ProQuest Dissertations Publishing. Retrieved from https://search.proquest.com/dissertations-theses/how-use-intuitions-philosophy/docview/304996059/se-2?accountid=14782

Turri, J. (2017). Compatibilism and Incompatibilism in Social Cognition. *Cognitive Science, 41*, 403-424. doi:10.1111/cogs.12372

Turri, J. (2017). Compatibilism can be natural. *Consciousness and Cognition, 51*, 68-81. doi:10.1016/j.concog.2017.01.018

Vierkant, T., Deutschländer, R., Sinnott-Armstrong, W., & Haynes, J.-D. (2019). Responsibility Without Freedom? Folk Judgements About Deliberate Actions. *Frontiers in Psychology*. doi:dx.doi.org/10.3389/fpsyg.2019.01133

Vihvelin, K. (2011). How to Think about the Free Will/Determinism Problem. In J. K. Campbell, J. Rheins, M. O'Rourke, & M. H. Slater (Eds.), *Carving Nature at Its Joints : Natural Kinds in Metaphysics and Science* (pp. 313-340). Cambridge: MIT Press.

Vihvelin, K. (2013). *Causes, Laws, and Free Will: Why Determinism Doesn't Matter.* Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199795185.001.0001

Vihvelin, K. (2015). How Not To Think about Free Will. *Journal of Cognition and Neuroethics, 3*(1), 393-403.

Viney, W., Waldman, D. A., & Barchilon, J. (1982). Attitudes Toward Punishment in Relation to Beliefs in Free Will and Determinism. *Human Relations, 35*(11), 939-950. doi:10.1177/001872678203501101

Wisniewski, D., Deutschlaender, R., Haynes, J.-D., & Jong, J. (2019). Free will beliefs are better predicted by dualism than determinism beliefs across different cultures. *PLoS ONE, 14*(9), e0221617. doi:dx.doi.org/10.1371/journal.pone.0221617

Woolfolk, R. L., Doris, J. M., & Darley, J. M. (2006). Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition, 100*(2), 283-301. doi:10.1016/j.cognition.2005.05.002