

UC Santa Cruz

UC Santa Cruz Electronic Theses and Dissertations

Title

Determinism, Alief, and Observer-Dependent Freedom: How to Mitigate the Consequences of Deterministic Thinking

Permalink

<https://escholarship.org/uc/item/64c328tm>

Author

Scherbart, Ryan

Publication Date

2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

**DETERMINISM, ALIEF, AND OBSERVER-DEPENDENT FREEDOM: HOW
TO MITIGATE THE CONSEQUENCES OF DETERMINISTIC THINKING**

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

DOCTOR OF PHILOSOPHY

in

PHILOSOPHY

by

Ryan Jon Scherbart

June 2013

The Dissertation of Ryan Scherbart
is approved:

Professor Jonathan Ellis, Chair

Professor Daniel Guevara

Professor John Bowin

Tyrus Miller
Vice Provost and Dean of Graduate Studies

Copyright © by

Ryan Scherbart

2013

Table of Contents

Abstract.....	v
Acknowledgements.....	vii
Introduction.....	1
 1. Determinism and Compatibilism.....	 3
1.1 Varieties of Determinism: Causal, Theological, Logical, and Bioenvironmental.....	4
1.2 Predictability and Determinism.....	11
1.3 Six Definitions of Scientific Determinism.....	12
1.4 Indeterminism and Quantum Mechanics.....	18
1.5 Chaos Theory.....	24
1.6 The Principle of Alternative Possibilities.....	27
1.7 Compatibilism Introduced.....	28
1.8 Classical Compatibilism: Hobbes.....	29
1.9 Conditional Compatibilism: Ayer.....	31
1.10 Contemporary Compatibilism: Frankfurt.....	36
1.11 Frankfurt's Account of Free Will.....	39
1.12 Objections to Frankfurt.....	41
1.13 Contemporary Compatibilism: Strawson.....	43
1.14 The Metaphysics of Strawson's Account.....	46
1.15 Problems for Strawson.....	50
1.16 Contemporary Compatibilism: Other Accounts.....	52
 2. Free Will and Alief.....	 60
2.1 An Increase in Deterministic Thinking.....	61
2.2 Determinism and Recent Work in Psychology.....	64
2.3 Baumeister et al. (2009): Belief in Determinism and Increased Aggression and Reduced Helping.....	67
2.4 Stillman and Baumeister (2010): Belief in Determinism and Reduced Learning from Guilt.	72
2.5 Stillman et al. (2010): Belief in Determinism and Reduced Satisfaction and Productivity in the Workplace.....	75
2.6 Mitigation of the Consequences: Convince Public Determinism is False or that Compatibilism is True.....	78
2.7 Mitigation of the Consequences: Inducing or Strengthening Alief in Free Will.....	83
2.8 Gendler's 4 Cases of Alief.....	84
2.9 Gendler's Characterization of Alief.....	90

2.10	Can Aliefs be Induced?	94
2.11	Ways to Induce Free Will Aliefs.....	96
2.12	The Appeal to Texts, Film, and Memory.....	97
2.13	Self-Talk.....	100
3.	Free Will₂ and the Law	108
3.1	Assumptions of the Criminal Law.....	108
3.2	Free Will ₂	109
3.3	Does the Law Need Amending?.....	114
3.4	Would Incorporating Free Will ₂ into the Law be Morally Justified?.....	121
3.5	Fischer, Ravizza, and Strawson.....	123
3.6	Wallace and Strawson.....	128
3.7	Southwood and the Arbitrariness Worry.....	132
3.8	Could Justice be Based on a Falsehood?.....	135
3.9	Similar Proposals?.....	137
3.10	Incorporation of Free Will ₂ into the Law.....	141
	Conclusion	145
	Bibliography	148

Abstract

DETERMINISM, ALIEF, AND OBSERVER-DEPENDENT FREEDOM: HOW TO MITIGATE THE CONSEQUENCES OF DETERMINISTIC THINKING

Ryan Scherbart

The problem of free will is one of the most intractable in the history of philosophy. Philosophers worry that there can be no free will in a determined world. The aim of this dissertation is not to attempt to solve this problem but to address two problems with deterministic *thinking*. A finding from recent work in psychology is that people with a strong belief in determinism or disbelief in free will tend to be more aggressive, less helpful, less likely to learn from guilt, and less productive and satisfied at work than those who do believe in free will (or don't believe in determinism). I will present several ways we might mitigate these effects, the most noteworthy of which involves an appeal to Tamar Gendler's notion of 'alief'. I argue the induction of an *alief*—or the strengthening of an already existent alief—in free will would mitigate the negative consequences of a strong belief in determinism or disbelief in free will. Following this, I investigate the problem of deterministic thinking with respect to the criminal law. It has been argued by Michelle Cotton, among others, that the criminal law (at least in the U.S.) proceeds on the assumption that to be convicted of a criminal offense the person in question could have done otherwise. Since the case has been made that no one could have done otherwise than she in fact did (because of determinism), legal conviction may not be possible. I

argue that incorporating into the law a concept I call *free will*₂—a concept dependent upon the judgments of a hypothetical observer—would make convictions legal *and* justified—or would at least close the legal loophole.

Acknowledgments

I first acknowledge my primary advisor, Jon Ellis, whose work and philosophical mind is inspiring. Our meetings over the past six years were invaluable. He made me a better philosopher. I thank him for the time we spent working together: in the office, over phone calls, and even via cross-country Skype sessions. Without Jon this dissertation would not be possible. He served as my cohort's pro-seminar leader; so he has been a philosopher's spirit-guide from the beginning. I give many thanks to Daniel Guevara, also, who shares my love of moral philosophy. Dr. Guevara always provided very helpful feedback and has been an advocate through the duration of the doctoral program. John Bowin, too, I must thank. Dr. Bowin made himself readily available and was a source of valuable comments, especially on early drafts of this dissertation. I must also mention David Hoy, without whom, I wouldn't be where I am today. I audited Dr. Hoy's History of Consciousness seminar in the Fall of 2006. He introduced me to some of the other UCSC grads. Without his recommendation and endorsement, I am sure my application to this Ph.D. program would have fallen to the wayside.

I am indebted to the members of the UC Santa Cruz Philosophy Department including, Michael Hicks (whom I owe a special thanks for introducing me to the work of Tamar Gendler), Jocelyn Hoy, Rasmus Winther, Ric Otte, Jorge Hankamer, Paul Roth, Janette Dinishak, and Abe Stone. I owe thanks to Mark Akesson, too, a member of the Biomolecular Engineering Department who served on my qualifying

exam committee. The ideas and arguments in this dissertation were fostered by conversations with my fellow philosophy graduate students. In particular, I thank Andy Delunas, Evan Lam, Kaija Mortensen, Lucas McGranahan, Amena Coronado, Alexis Mourenza, Lourdes Ortiz-Bautista, Ben Roome, Javier Cardoza-Con, Kyle Roberson, Jake Metcalf, and Daniel Pilchman. Other members of the UC Santa Cruz community I thank are Brij Lunine, Amy Weaver, and Lynn Galiste.

Certain portions of this dissertation were presented at a works in progress colloquium at UC Santa Cruz in Spring 2010 entitled, “Can Response-Dependence Save Free Will?”

Many thanks to all.

Santa Cruz, June 2013

Dedication

For my family—past, present, and future.

Introduction

This dissertation will address growing concerns about ‘deterministic thinking’ that has made its way into the minds of the general public and the criminal law.

There is reason to think an increasing number of people believe in determinism and disbelieve in free will.¹ This is cause for concern, I will argue, because recent findings from psychology indicate a belief in determinism or disbelief in free will has negative social implications, including increased aggression, decreased helpfulness, decreased learning from guilt, and decreased satisfaction and productivity in the workplace. The main thesis of the second chapter of this dissertation is that inducing an ‘alief’² in free will would serve to mitigate these negative consequences. In that chapter, I will also discuss other ways of mitigating the consequences, including the encouragement of self-talk monitoring and the reflecting on of texts, films, and memories I describe as ‘freedom inspiring’. In the third chapter, I will discuss the implications of deterministic thinking in the criminal law. Michelle Cotton has argued, among others, that the criminal law (at least in the U.S.) proceeds on the assumption that any person who is convicted of a criminal offense could have done otherwise (than commit the offense). Assuming no one could have done otherwise than they, in fact, did (because of causal determinism), legal conviction may not be possible. The main thesis of the third and final chapter, then, is that incorporating

¹ The details of the evidence and precise definitions of ‘determinism’ and ‘free will’ to be offered.

² According to Gendler (2008a), p. 642, an alief is “a mental state with associatively-linked content that is representational, affective and behavioral, and that is activated—consciously or nonconsciously—by features of the subject’s internal or ambient environment.”

into the law a concept I dub *free will*₂—a concept dependent upon the judgments of a hypothetical observer—would allow for convictions that are legal *and* justified. The incorporation of *free will*₂ into the law would at least close the legal loophole, I will argue.

The first chapter will be expository only. There, I will provide a literature review with respect to two concepts central to the subsequent chapters, namely *determinism* and *compatibilism*. The former is central because it is deterministic thinking that allegedly is problematic for social behavior and the criminal law. The latter will be important for two reasons. Firstly, exposing the public to cutting-edge compatibilist accounts will be one of the ways I explore as a means of mitigating the consequences of a strong belief in determinism (or disbelief in free will). Secondly, the account of *free will*₂ I offer in the third chapter echoes the compatibilist account of P.F. Strawson, whose position I will cover in detail in the second section of chapter one.

1. Determinism and Compatibilism

Recent work in psychology has indicated a strong belief in determinism or weak belief in free will has negative social consequences. Vohs & Baumeister (2009) tie a disbelief in free will to cheating, stealing, aggression, and reduced helping. Stillman and Baumeister (2009) suggest that a strong belief in determinism leads to reduced learning from emotions like guilt. And Stillman et al. (2010) show that subjects with weak beliefs in free will were less satisfied at work and less productive. These are three of a growing number of studies and experiments. In addition to sobering implications, these results suggest many people are incompatibilists, i.e., people who maintain that free will cannot exist in a deterministic world.

In the *second* chapter of this dissertation, I will attempt to show that these studies are really something philosophers and the general public should be worried about. The main task of chapter two, then, is to investigate ways to mitigate these negative social ramifications. In the final chapter, I advance the thesis that because of assumptions in the law about free will and the growing acceptance of determinism, courts would need to revise their definitions of ‘guilty’ and ‘free will’ to properly function.

This chapter will be entirely expository. I will provide background for two philosophical concepts that will be important in the proceeding chapters: *determinism* and *compatibilism*. One strategy I present in the next chapter to mitigate the negative consequences is to convince the public there is a viable variety of compatibilism or that determinism is false (or that there’s a reason to think so). While there are many

important versions of compatibilism, it is beyond my purpose here to address them all. I will, however, focus on several contemporary accounts that have garnered much attention in the literature, paying special attention to the accounts from Harry Frankfurt and P.F. Strawson. In the second half of this chapter, I investigate both. That section will also include highlights from the history of compatibilism—especially the ones that set the stage for Frankfurt and Strawson. I will begin the second section by motivating the larger philosophical problem Frankfurt and Strawson both address—which is arguably the deeper underlying issue behind the alarming results from the psychologists. Posed as a question: how can we have free will in a world that is determined? So, the first section of this chapter will discuss what determinism is and why we should be worried about it.

1.1 Varieties of Determinism: Causal, Theological, Logical, and Bioenvironmental

It is an understatement to say there are many varieties of determinism. Jordan Sobel claims there are at least ninety.³ When asked if determinism is compatible with free will, he quips, “the answer... can be only, Yes and No. Everything depends upon what one means by ‘determinism,’ and by ‘free will.’”⁴ It is beyond the scope of this chapter to touch upon every kind of determinism. I will instead focus on *causal determinism*, as is it is considered by many to be the greatest threat to free will.⁵ But

³ This claim about Sobel comes from Bishop (2002), p. 111.

⁴ Sobel (1998), p. 77.

⁵ Of course, it’s not *the theory* that threatens free will; it’s the implication from *the truth* of the theory.

first I will distinguish causal determinism from other major varieties. These other varieties may be *consistent* with causal determinism but, as we will see, have important differences. I will then explain the relationship between causal determinism and predictability and proceed to walk through several versions of causal determinism itself. Following that, I discuss how work in quantum mechanics has affected the acceptance of causal determinism. Even if *universal* causal determinism is false, there may be *weaker* forms of determinism, the truth of which would threaten free will. Along the way, I will talk about why causal determinism—including these weaker forms—are worrisome for proponents of free will and how compatibilists have attempted to show that determinism and free will are consistent.⁶ This discussion will segue into §2, the section devoted to compatibilism. By this transition, it will become clear why certain versions of causal determinism have widespread support, and how this has motivated the problem of free will.

Causal determinism is distinct from what I take to be three other major varieties of determinism, that is, *theological*, *logical*, and *bio-environmental*. Some philosophers make a distinction between *causal* determinism and *scientific* or *physical* determinism. Most agree that scientific or physical determinism is a *sort of* causal determinism; however, scientific determinism must reference laws of nature, while causal determinism need not. For instance, if you claim all events are caused by sufficient antecedent conditions, then Eshleman (2009) and the like would count you as a causal determinist, but not a scientific determinist because you make no

⁶ Or more precisely that the *proposition* (a): *The world is determined* is consistent with *proposition* (b): *Some human actions are free*.

mention of laws. Theories of causal determinism appeared in the work of the presocratics but substantially gained momentum in the wake of Newton's discoveries. The kind of determinism I'm primarily concerned with here is—on Eshleman's definition—scientific, but I will use the broader, more ubiquitous term, 'causal determinism'.

As Van Inwagen puts it, causal determinism is the view that “the past determines a unique future (given the past and the laws of nature, the future is determined in every detail).”⁷ In a causally determined world, the future is *fixed*. When I say ‘the future is fixed’, I mean the future is such that it will unfold in a certain way and can't unfold in any other way. In the case of *causal* determinism, the future is fixed in virtue of the laws of nature in conjunction with prior conditions. Maudlin (2007) sees the laws of nature as ‘pushy explainers’, which make things unfold in particular ways. On this account even the actions we consider to be free are part of nature's unfolding. Our lives are part of a causal chain fused by natural laws, rendering all our actions unavoidable. As we shall see, this unavoidability is at the heart of the free will debate.

In the case of *theological determinism*, fixedness of the future is entailed by (though not necessarily *caused* by) the omniscience—or more specifically the divine foreknowledge—of God.⁸ If God is omniscient, then God knows the future—including all the future acts of humans. Proponents of theological determinism posit

⁷ Van Inwagen (1983), p. 342.

⁸ This is not to say God's divine foreknowledge *does* the fixing. A theological determinist could very well hold that the future is fixed because of the law of nature in conjunction with prior conditions.

if God knows that person *P* will do *X* at time *t*, then the proposition, *person P will do X at time t*, is true. Suppose God knows that I, Ryan, will eat a grilled cheese sandwich on October 21st, 2015. Thus according to these proponents, the proposition, *Ryan Jon Scherbart will eat a grilled cheese sandwich on October 21st, 2015*, must be true. Moreover, it would be impossible for me to do otherwise—impossible for me to *not* eat a grilled cheese on that day. For me to do otherwise would entail the falseness of the aforementioned proposition, entailing God was wrong about the future, in effect denying God's perfect knowledge. But perfect foreknowledge is the primary assumption of theological determinism. It should be noted, too, that I'm contrasting theological determinism from causal determinism not because they are inconsistent. The two varieties are consistent, since the world might be such that causal determinism is true and there exists a being with perfect foreknowledge. But they are distinct views, and one of them could be true while the other false. For example, the world could have been such that there were no causal laws and events that unfold randomly, but such that God's foreknowledge entailed true propositions about the future. Suppose, in this chaotic world God knows I will become a teacher. In that world my teaching career is unavoidable, but this would have nothing to do with the laws of nature.⁹

On a *logically deterministic* account, the source of the problem is a principle of logic, namely the law of bivalence, under which every proposition is either true or false. If we take the law of bivalence to hold for *future tense* propositions, then the

⁹ For more on theological determinism see Zagzebski (1985), Luis de Molina (1988), and Warfield (1997).

future is fixed by the present. The following is an argument for logical determinism (sometimes called ‘logical fatalism’) put forth by Barnes and Cameron (2009). It is an alteration of an argument Aristotle makes in *De Interpretatione*, chapter nine.

- (1) Either it’s true that there will be a sea battle tomorrow or it’s true that there won’t be a sea battle tomorrow.
- (2) If it’s true that there will be a sea battle tomorrow, then it’s true now that there will be a sea battle tomorrow, and likewise, *mutatis mutandis*, if it’s true that there won’t be a sea battle tomorrow.
- (3) If it’s true now that there will be a sea battle tomorrow, or true now that there won’t be, then how tomorrow is (at least with respect to sea battles) is settled by how the present is.
- (4) Therefore, how tomorrow is (at least with respect to sea battles) is settled by how the present is.
- (5) Since we were dealing with an arbitrary event at an arbitrary future time, how the future is in all respects is settled by how the present is.¹⁰

In general, the view is if it’s *presently* true that so-and-so will happen in the future, then that future is unavoidable. Logical determinism, too, is consistent with causal determinism, for it could be the world is causally determined and that the law of bivalence holds for future tense propositions, rendering the future unavoidable. But again, logical and causal determinism need not both be true. For instance, it could be the case the world is chaotic and causal determinism is false but that the law of bivalence applies to future tense propositions, thereby settling the future on the basis of the present.¹¹

Bio-environmental determinism is a broad variety, on which an individual’s behavior may not be fully causally determined but is at least substantially influenced by (a) her biology, i.e., her genes, DNA, epigenome, neurology, psychology etc.—

¹⁰ Barnes and Cameron (2009), p. 292.

¹¹ For more on logical determinism, see Taylor (1962).

and or (b) her environment, i.e., her culture, socio-economic status, upbringing, peers, and so on. I count this as a single variety because of the inherent interaction between biology and environment. Consider a grilled cheese sandwich case for a bio-environmental determinist. On her view, I might be able to avoid making the sandwich, but my biology and/or environment might *push* me toward it. I liken this *push* to the influence from peer pressure. If Tommy—a fourteen-year-old boy—is being pressured by his peers to smoke a cigarette, we can imagine him resisting, but doing so may require a tremendous effort on his part. Note also that this sort of determinism only affects the actions of beings like humans and not all events in the world. Further, some may hesitate from counting this as a variety of determinism because it is consistent with indeterminism (I discuss this later).

Bio-environmental determinism—like theological and logical determinism—is consistent with causal determinism. For it could be that bio-environmental determinism is merely *part* of the causal determinist's story: one might envision our biology and environment as two of the cogs in the deterministic universe. If that were true, then in the case of Tommy, not only does smoking *feel* irresistible to him, it—in effect—*is*. Tommy can do nothing but smoke the cigarette on account of his biology and environment. These factors determine his behavior, rendering it unavoidable. This concludes my discussion of non-causal forms of determinism. Henceforth, I will often drop the 'causal' from 'causal determinism'.¹²

¹² For more on bio-environmental determinism, see Peet (1985) and Dingwall, et al. (2003).

Many philosophers trace determinism back to the presocratic atomists, namely, Leucippus and his student Democritus, for they offer probably the earliest version of it. If the world consists of nothing but atoms in a void governed by causal laws—as they both held—then “nothing occurs at random, but everything for a reason and by necessity.”¹³ But it was because of the work of Newton that determinism gained force, for Newton’s laws were empirically verified, expressed mathematically, and ubiquitous. Pierre-Simon Laplace is often credited as the first philosopher to write on the consequences of this new worldview. He expressed them in his famous 1820 narrative about a knowledgeable demon:

We ought to regard the present state of the universe as the effect of its antecedent state and as the cause of the state that is to follow. An intelligence knowing all the forces acting in nature at a given instant, as well as the momentary positions of all things in the universe, would be able to comprehend in one single formula the motions of the largest bodies as well as the lightest atoms in the world, provided that its intellect were sufficiently powerful to subject all data to analysis; to it nothing would be uncertain, the future as well as the past would be present to its eyes.¹⁴

However, many claim while Laplace beautifully captures the predictability that falls out of determinism,¹⁵ he misleads the reader into thinking that determinism and the theory that all future events are predictable are equivalent, which is false. Many have pointed out this mistake, including Macintyre, who says, “The question of predictability and the question of causality may be connected but are certainly

¹³ Translated from “οὐδὲν χρῆμα μάτην γίνεται, ἀλλὰ πάντα ἐκ λόγου τε καὶ ὑπ’ ἀνάγκης” from Leucippus, Fragment 569 - from Fr. 2 Actius I, 25, 4.

¹⁴ Laplace (1820), p. 4.

¹⁵ NB: Some determined worlds are fully predictable but some aren’t, and some fully predictable worlds are determined and some aren’t. In other words, it’s false that a world is determined if and only if it is fully predictable.

distinct.”¹⁶ Laplace wasn’t the only great thinker to make the mistake; Popper was guilty of the same conflation.¹⁷

1.2 Predictability and Determinism

But probably the reason why Laplace’s story is still relevant today is it poignantly brings to light our anxiety about determinism. Its tone puts a chill down the back of anyone who puts stock in her own free will. If I tell someone all her future acts—little ones, like choosing what to eat for lunch and big ones, like deciding which career to pursue or whom to marry—are predictable, then it is hardly likely she would consider those acts free. A common belief is a fully predictable life is one without alternative possibilities, and a life without alternative possibilities is one without free will. Later, I will discuss whether this belief is ill-founded, but for now let us recognize it as a common intuition.

We might be tempted to say if a theory of determinism doesn’t entail full predictability, then the theory shouldn’t be considered deterministic. However, Hoefer (2010) claims it is perfectly sensible for a deterministic world to not be fully predictable. Most physicists would say we live in a deterministic universe, yet “19th and 20th century mathematical studies have shown convincingly that neither a

¹⁶ Macintyre (1957), p. 30.

¹⁷ For more on the relationship between determinism and the theory that all future events are predictable, see Boyd (1972).

finite, nor an infinite but embedded-in-the-world intelligence can have the computing power necessary to predict the actual future, in any world remotely like ours.”¹⁸

So—to be clear—(a) determinism is distinct from the theory that all future events are predictable and (b) even a non-predictable, deterministic world is problematic for free will—especially if *the ability to do otherwise than we, in fact, do* is taken as a necessary condition for freedom.¹⁹ A non-predictable deterministic world is troublesome for free will because the unavoidability of a unique future still exists “whether [or not] any demon (or even God) can, or cares to, actually predict what we will do.”²⁰ It is “the strings of *physical necessity*, linked to far-past states” that determine all future events.²¹ The compatibilist’s job, then, is to either show that even with such physical necessity we have the ability to do otherwise or that we can have free will without it. I will look at some historical attempts that attempt to show just that as well as objections that suggest the attempts end in failure.

1.3 Six Definitions of Scientific Determinism

Thus far I have said a lot about what determinism *isn’t*, but I haven’t said much about what it *is*. I have offered one definition by Van Inwagen. Call it D1:

- (D1) The past determines a unique future (given the past and the laws of nature, the future is determined in every detail).

¹⁸ Hoefer (2010).

¹⁹ I will sometimes shorten ‘the ability to do otherwise than we, in fact, do’ to ‘the ability to do otherwise.’

²⁰ Hoefer (2010).

²¹ *Ibid.*

I began with this analysis because it is clear and succinct. It is circular, however, because in defining ‘determinism’ it uses the words, ‘determines’ and ‘determined’. Let us consider other definitions of ‘determinism’, then. These versions—which I will introduce in just a bit—are the most worrisome in the literature. This is because the truth of any one of them poses a threat to free will, and there is substantial evidence for them (or at least one), leaving little room for free will. Theological and logical determinism threaten free will, too, but the former requires the existence of God (or at least a being with perfect foreknowledge) and the latter faces serious objections, which I shall not pursue here. Bio-environmental determinism is also worrisome, but has yet to be established to the same degree as the other varieties and—as discussed before—may collapse into causal determinism—or at least may depend on its truth. Showing that we do have free will despite the truth of determinism is the challenge taken up by the compatibilist, which I will discuss in detail in the second half of this chapter. In that section, I will discuss the compatibilism of Frankfurt and Strawson, both of whom subscribe to Van Inwagen’s view—or something very close to it.

I’d like to quickly say one more thing before we compare D1 to its ‘siblings’. It is a description of determinism (also by Van Inwagen) that will be helpful to keep in mind before we proceed. In addition to D1, Van Inwagen maintains that “determinism is the thesis that if time could be ‘rolled back’ to any past instant, and then allowed to ‘go forward again,’ then there is no question but what history would ‘repeat’ itself: we could be certain that things would happen ‘again’ just as they

happened the ‘first time.’”²² I don’t know if this is so much a definition as it is an apt rhetorical device, but either way it produces a helpful image to keep in mind as we move forward.

The common thread among D1 and its siblings is they all mention a relationship between (a) conditions or states of affairs and (b) laws of nature. The relationship is such that the conjunction of the two establishes future conditions or states must happen. Here is the second of six such analyses:

- (D2) The state of the universe at any time is wholly and unequivocally determined by the state of the universe at prior times and the physical laws of nature.²³

In D2 Hodgson qualifies the laws of nature as *physical*. I think this qualification is implied by Van Inwagen in D1, as it has been and will be my implication throughout this chapter. Hence, this definition from Hodgson I take not to be much different from D1. One difference, however, is that Van Inwagen’s definition contains an ambiguity that D2 does not. The ambiguity stems from the terms ‘the past’ and ‘the future’. It’s not clear on D1, if ‘the past’ is meant to refer to a single moment in the past or to the whole past—which would include every past moment. Likewise, it’s unclear if ‘the future’ is meant to refer to a single moment in the entire future or the entire future—also which would include every future moment. I will assume that Van Inwagen was being loose in D1, but meant something like D2. On D2, the state of the universe at T_1 in conjunction with the laws of nature would necessitate a state at some future time, T_2 .

²² Van Inwagen (1974), p. 9.

²³ Hodgson (2005), p. 85.

Now consider D3:

- (D3) Everything in the universe is in some sense physical, and every event is rendered inevitable by virtue of the past states of this physical universe in conjunction with causal processes governed by laws of nature.²⁴

In D3, Pereboom mentions ‘causal processes’, which doesn’t appear in the analyses from Van Inwagen and Hodgson. However, I think causal processes are implied by D1 and D2 because both reference laws, which are inherently causal. So I take that difference to be only a superficial one. In addition, Pereboom uses the term ‘inevitable’, which doesn’t appear in D1 or D2. But I don’t think this makes his analysis any different either. The states determined on D1 are inevitable. The same goes for D2. Early in the chapter, I talked about events being ‘unavoidable’, but I could have just as easily used ‘inevitable’. If D3 is true, everything in the universe is physical. I assume proponents of D1 and D2 also make this assumption. I won’t go as far to say one must endorse physicalism to be a determinist, but I acknowledge determinists typically are physicalists. One final note about D3 is unlike the previous definitions it doesn’t include the word, ‘determined’, in the definition of ‘determinism’.

Let us now turn to D4:

- (D4) A past event E is ‘totally determined’ if E is subsumable under a universal law of nature; that is the occurrence of E was deducible from a description of the conditions that obtained before its occurrence and certain universal laws.²⁵

²⁴ Pereboom (2006), p. 467.

²⁵ Feinberg & Shafer-Landau (2008), p. 408.

D4 isn't actually a definition of determinism because it doesn't claim any events are determined. However, I include it in this list as it makes explicit an important concept we don't explicitly see in D1 through D3, namely the concept of *deducibility*. I do however take D1-D3 to imply the deducibility claim in D4.

Now consider D5:

- (D5) Every true proposition (E at t) to the effect that an event of kind E has taken place, is taking place, or will take place at some time t is related to some true proposition (C at t') to the effect that an event of some kind C takes place at some earlier 'time' t', and to a true law of nature L, such that the conjunction of (C at t') and L entails (E at t).²⁶

Sobel defines determinism in terms of true propositions, instead of descriptions as Feinberg and Shafer-Landau do. We can think of determinism, on Sobel's account, as a relation of event-pairs: event-pair, C at t', and event-pair, E at t. Both event-pairs are bound by the laws of nature, L, such that C at t' together with the laws necessitates that E will occur at t (or that the proposition describing E at t will be true). The deducibility we saw in D4 can be found here, too. The terms 'valid' and 'truth preservation' are suitable. If E at t is true, then some C at t' must be true. Why is this the case? Because of the laws of nature, L. They are the 'pushy explainers'. Given prior conditions and the laws, the future is inevitable.

I conclude this list with D6, also from Van Inwagen, as it is the most precise. Van Inwagen claims he needs three subordinate notions to adequately define determinism: the notion of (a) *a proposition*, (b) the notion of *the state of the entire*

²⁶ Sobel (1998), p. 84.

physical world at an instant, and (c) the notion of *a law of nature*. Propositions for Van Inwagen are “non-linguistic bearers of truth-value;” however, beyond that it is unlikely to matter for D6 which account of propositions one subscribes to as long as it includes their usual features, e.g., “they are either true or false; the conjunction of a true and a false proposition is a false proposition; they obey the law of contraposition with respect to entailment.”²⁷

There are many accounts of ‘the state of world’, that could be used in Van Inwagen’s definition of determinism, but they must include at least two conditions: (C1) nothing about states at other times must follow *logically* from the state of the world at a particular time. (C2) Any observable change in the way things are must entail a change in the state of the world.²⁸

Van Inwagen’s theory of determinism (call this one D6) is the conjunction of the following two theses:

(T1) For every instant in time, there is a proposition that expresses the state of the world at that instant.

(T2) If *A* and *B* are any propositions that express the state of the world at some instants, then the conjunction of *A* with the laws of physics entails *B*.²⁹

With the thesis spelled out, it is clear, Van Inwagen notes, why the first condition about states of the world—(C1)—is important. Without it, we could “build sufficient information about the past and future” into *A* and *B*, such that *A* alone, i.e., without conjoining it with the description of the laws of nature, would entail *B*. But if this

²⁷ Van Inwagen (1975), p. 185.

²⁸ *Ibid.*, p. 186.

²⁹ *Ibid.* This assumes *A* is prior to *B*.

were the case determinism would be “a mere tautology”. From this point forward, when I use the term, ‘determinism’, I mean D6.³⁰

1.4 Indeterminism and Quantum Mechanics

But is determinism *true*? I’ve talked about what causal determinism *is not* and what it *is*, but now I ask, is it true? On a Newtonian framework, the answer is unequivocally, *yes*. Newton and his disciples believed in determinism because Newton’s findings provided overwhelming evidence for physical laws, and these laws were such that conditions in conjunction with them would necessitate future conditions. The experiments Newton conducted were *repeatable*. The repeatability of classic experiments in physics is in the spirit of the second description of determinism from Van Inwagen that I included—that if we could roll back time history would repeat itself. However, because of the work in quantum mechanics beginning in the 1920s, the Newtonian picture has become murky. In fact, most experts in relevant fields will explain that determinism in a universal sense (i.e., on a micro *and* macroscopic level) is false. If they are correct, what does this mean for free will? I will discuss the answer to this and related questions now.

If determinism is false, then some form of *indeterminism* is true. Indeterminism is the theory that there is some part or level of the universe that is not determined. Most who work in quantum mechanics claim that many events on the

³⁰ Van Inwagen (1975), p. 186.

quantum or *micro-level* world are indeterminate. This means there are some events that are not necessitated by past events. Since determinism is so problematic for free will, there are those who believe any chance we have for freedom will stem from quantum indeterminacy. But this may be a false hope because as Macintyre puts it, “If we could show that some events were uncaused we should have done nothing to the point in our present dilemma [i.e., the dilemma of determinism]. For to say that any given event is uncaused is surely to say that such an event is random. What is random is no more free than what is caused.”³¹ In other words, if ‘free’ human actions are rooted in quantum indeterminacy, then, according to Feinberg and Shafer-Landau (2008), they are “arbitrary and unintelligible.”³² Proponents of free will seem to face a dilemma with two horns. If ‘free’ actions are determined, then they aren’t really free, incompatibilists say. If ‘free’ actions are not determined, then they aren’t really free either, Macintyre argues. Moreover, there are those like Strawson who claim free actions are *required* to be determined.³³

But let’s set that issue aside. Let us discuss the details of indeterminism. Pereboom argues that “on the standard interpretation of quantum mechanics... the physical world is not in fact deterministic, but is rather governed by probabilistic statistical laws.”³⁴ This *isn’t* a controversial notion either, in fact, “it’s pretty widely

³¹ Macintyre (1957), p. 30.

³² p. 409.

³³ Strawson (1962). Here Strawson talks about the concepts of *moral obligation* and *responsibility*, but to any reader of his paper it is clear that he would extend these words to the concept of *freedom*. So his claim is that some philosophers maintain that the truth of determinism is *required* if we are to make sense of the concepts of moral obligation, responsibility, *and* free will.

³⁴ Pereboom (2007), p. 112. By the ‘standard interpretation’, Pereboom means the Copenhagen interpretation. There are many interpretations of quantum mechanics. On one advanced by Everett

accepted that we don't right now have any good arguments for universal determinism... [because] we have no good arguments for micro-level determinism, because quantum mechanics contains probabilistic laws, and there are no good arguments for any deterministic interpretation of these laws."³⁵ These claims make clear the near consensus on this issue: the notion of universal determinism that came out of the Newtonian world is mistaken. Notice, however, there are still *laws* that hold on the micro-level. But what is the nature of these laws and how does this entail the quantum world is not determined? For the answer, let's turn to the work of physicists from the 1920s. That work tells us the behavior of some particles in atoms is not determined by prior conditions in conjunction with laws like behavior in the macroscopic world. For instance, if an electron's spin is up in the x-direction, then the spin in the y direction—a direction *orthogonal* or at a right angle to the x-direction—is in what's called a *superposition*. The superposition, in this case, is "a kind of indeterminate state, a sort of 'hovering between' spin-up and spin-down."³⁶ Thus, when the electron is measured for spin in the y-direction, there is a 50% chance its spin will be up and a 50% chance the spin will be down. It's important to note that it is not the case that the spin of the electron in the superposition is up or down *and we just don't know it*. Because of the work of Kochen and Specker (1967), there is strong evidence that the superposition state is "physically real, and the measurement

(1973), quantum events, like macro-level events, obey deterministic laws and represent branch points separating distinct branches of the universe. For instance, on one branch an electron's spin may be up while on another it is down. On this view, all branches of the universe are equally real, yet cannot interact with one another.

³⁵ Balaguer (2009), pp. 6-7.

³⁶ *Ibid.*, p. 12.

process involves a *collapse*—what’s known as a collapse of the wave function—into a spin-up or a spin-down state.”³⁷ The most widely accepted interpretation of quantum mechanics is the Copenhagen interpretation, which we owe to the contributions of Niels Bohr. Under the Copenhagen interpretation, the lack of certainty in quantum mechanics is not something scientists could overcome with enough time and study. The behavior of the phenomena in the quantum world, moreover, is inherently indeterministic, such as electronic spin (as mentioned) and the duality of wave-particles.

One of the major quantum world surprises is that the collapse of the wave function is due to observation or measurement of the electron. This is known as *the observer effect*. Experiments in *classical* physics can be altered by this phenomenon too, but properly done these experiments can eliminate the effect. But eliminating the observer effect in the quantum world is impossible. The observer effect is often confused with Heisenberg’s uncertainty principle. Both are relevant to quantum indeterminacy, but they should not be conflated. Heisenberg theorized that if you knew the momentum of an electron, its position would be indeterminate, and if you knew with certainty its position, its momentum would likewise be indeterminate. Moreover, the more accurately you can determine the position of the electron, the less confident you will be with respect to its velocity *and vice versa*.

Suppose our best theories of quantum mechanics are correct, and because determinism does not hold in the quantum world, universal determinism is false.

³⁷ Balaguer (2009), p. 13. NB: On the many worlds interpretation (see Everett (1973)), no observer is required to collapse the wave function.

What does this mean for free will? For thinkers whose theories require the ability to do otherwise for free will, quantum indeterminacy will only be interesting if it makes that ability possible. But this doesn't seem to be the case. Although the micro-level world is indeterminate, it does not follow that the macro-level world is. That is, quantum indeterminism may negate universal determinism, but it doesn't necessarily rule out *macro-level* determinism. In fact, many physicists in quantum mechanics maintain quantum indeterminacy is irrelevant with respect to macro-level events *and this includes the neural events in the human brain*. If indeterminacy is the only hope for free will, then many say that quantum indeterminacy must somehow 'spill over' into the world of human behavior, including brain activity. But the spill over theory isn't well supported. Many, like Hodgson (2005), claim that for sufficiently large phenomena (like billiard balls and rockets), the indeterminism of quantum mechanics is moot. He states:

For most practical applications, the inability of QM [quantum mechanics] to do more than give probabilities of measurements does not matter much, since the indeterminacies and indeterminism of QM are generally at atomic scales, and in fact QM both gives virtual certainties for the behavior of systems comprising large numbers of particles of matter or radiation, and also confirms the substantial accuracy of classical physics for macroscopic systems.³⁸

Or consider Bishop, who maintains that "quantum effects, though substantial when focusing on single atoms are presumed negligible when focusing on systems involving large numbers of atoms."³⁹ And *the reason* quantum indeterminacy doesn't

³⁸ Hodgson (2005), p. 91.

³⁹ Bishop (2002), p. 119.

infiltrate the macro-level events that comprise human neurology and behavior is a phenomenon known as *quantum decoherence*.

The idea behind quantum decoherence is essentially this: the indeterminacy of a micro-level event taken *individually* is significant, but in a macro-level event, e.g., the behavior of a neuron, the sheer number of micro-level events is so large that the indeterminacies average out to near certainty. Consider a coin toss. Toss it once and it's hard to predict what will happen: it might be heads; it might be tails. Toss it ten times, and it's still hard to predict what will happen. Five heads and five tails would be a good prediction. But as experience tells us, we could easily get seven heads and three tails. But the odds of getting 70% heads and 30% tails will significantly diminish if we substantially increase the number of coin tosses. Increase them to 10,000 and we'll see something close to 5,000 heads and 5,000 tails. With a million (or a billion!) tosses, the ratio approaches 50:50. Thus, in the macro-level world, the predictions of classical physics become virtually perfect, and this is true for the events relevant to free will, like neural events. Work on decoherence took form in the 1980s and is still fairly controversial. That said, "based on a calculation of neural decoherence rates," Tegmark, et al. (2000) "argue that the degrees of freedom of the human brain that relate to cognitive processes should be thought of as a *classical* rather than quantum system [emphasis mine]."⁴⁰ There is more to quantum

⁴⁰ p. 4194.

decoherence than what I have just described, but it's beyond my focus here to include more than this thumbnail sketch.⁴¹

1.5 Chaos Theory

Some argue that chaos theory might reveal an *amplification* of quantum indeterminacy, thereby countering quantum decoherence. Chaos theorists who specialize in classical physics have shown, for instance, that slight variations in initial conditions can drastically alter the outcome of events. For instance, if you put a coffee can in the middle of a pool table and take shots at it with the cue ball, you would notice dramatically different trajectories on even slightly different shots. This would mostly be the result of the convex surface's effect on the trajectory of the ball. The take home point of chaos theory is that seemingly miniscule changes in a system, over the course of a very short period of time, can yield substantial, wide-scale effects. Could chaos amplify the minute indeterminacy of the micro-level? Bishop argues:

Chaos could amplify quantum events, causing a single neuron to fire that would not have fired otherwise... this additional neural firing, small as it is, would then be further amplified to the point where the brain states would evolve differently than if the neuron had not fired. In turn these altered neural firings and brain states would carry forward such quantum effects that affect the outcomes of human choices.⁴²

So just like the smallest directional change of a pool shot could result in a very different trajectory and final resting position of a billiard ball, some argue quantum

⁴¹ For more on this, see Penrose (1989), Penrose (1997), and Tegmark, et al. (2000),

⁴² Bishop (2002), p. 119.

indeterminacy could be amplified in a chaotic system, such that a neuron might fire that might not have otherwise, which could result in an altered neurology and ultimately an altered will. But this analogy is very controversial. For one, it may be the human brain is not a chaotic system.⁴³ And secondly, while quantum indeterminacy may be sensitive to chaos *in theory*, “applying such arguments to *concrete* physical systems shows that the amplification process *may be severely constrained* [emphasis mine].”⁴⁴ That is, there are plausible models of a chaotic brain, but those models might not account for sophistication in the actual brain that may inhibit chaos amplification. That being said, there are members of the scientific community who endorse a chaotic model of the brain, e.g., Freeman & Skarda (1987) and Korn and Faure (2003), as well as work from philosophers, e.g., Kane (2007), who claims chaos amplification in the brain is the key to human freedom.

Let’s recap. Determinism negates our ability to do otherwise. Many stakeholders in the free will debate define free will in terms of the ability to do otherwise, and if they’re right, then determinism rules out free will. However, determinism—in a universal sense—is false because of quantum indeterminism. Perhaps quantum indeterminism plays some role in the actions we take to be free—Popper, Kane, and others have held this view. But there seems to be a consensus among quantum physicists that the indeterminacy of the micro-level does not affect ‘the world we live in’, i.e., the macro-level world. In other words, the macro world—including human behavior—is determined, at least virtually. However, maybe those

⁴³ See Diesmann, Gewaltig, and Aertsen (1999).

⁴⁴ Bishop (2002), p. 120.

physicists are wrong; maybe quantum indeterminism *does* play a role in the actions we consider free. How? Perhaps the human brain is a chaotic system, and quantum indeterminacy is *amplified* by this chaos. But even if *that* could be shown, it's still not clear how an indeterminate neurology would give or allow us the ability to do otherwise. Instead of this ability, a non-determined brain seems only to entail randomness. Recall Macintyre's position: "To say that any given event is uncaused is surely to say that such an event is random. What is random is no more free than what is caused."⁴⁵

So indeterminism, unfortunately, seems to be just as much a dead end as determinism. However, this may not mean the end for free will. I've said repeatedly that many of those who work on free will define it in terms of the ability to do otherwise. What makes determinism so lethal for many is it entails we don't have this ability. There is evidence that determinism is true—at least a *virtual* macro-level or *adequate* determinism, which would govern human decision making and behavior. If we accept this, then it seems we do not have the ability to do otherwise than we, in fact, do. The argument from the incompatibilist, who claims free will can't exist in an adequately determined world, might best be expressed as follows:

- (1) If adequate determinism is true, then no one ever has the ability to do otherwise than she, in fact, does.
- (2) If no one ever has the ability to do otherwise than she, in fact, does, then no one has free will.
- ∴ (3) If adequate determinism is true, then no one has free will.

⁴⁵ Macintyre (1957), p. 30.

The defender of free will seems to be backed into a corner. But there are those who have diligently tried to fight their way out, most notably Harry Frankfurt. Frankfurt does *not* define free will in terms of the ability to do otherwise. He maintains we can have freedom without it. And if that's the case then we can have our cake and eat it too. We can accept determinism and simultaneously hold that we have free will. That, in short, is compatibilism.

1.6 The Principle of Alternative Possibilities

Defining free will in terms of the ability to do otherwise had been the status quo until the 1960s. The ability to do otherwise is at the heart of what Frankfurt calls the principle of alternative possibilities (PAP).

PAP A person freely performed some action *A* only if she could have done otherwise (i.e., she could have avoided doing *A*).⁴⁶

But if adequate determinism and the PAP are true, free will seems impossible or at least very unlikely because adequate determinism means it's unlikely we have the ability to do otherwise. A.J. Ayer describes this problem quite succinctly in his opening to "Freedom and Necessity":

When I am said to have done something of my own free will it is implied that I could have acted otherwise; and it is only when it is believed that I could have acted otherwise that I am held to be morally responsible for what I have done. For a man is not thought to be

⁴⁶ Frankfurt frames the PAP in terms of moral responsibility instead of free will, but I will frame it in terms of free will, which Frankfurt, I believe, would endorse since he and many others hold that if someone is morally responsible for having done something, then she freely did it. When I say 'she freely did it', I mean she had free will in doing it. Also, I will sometimes use the term 'freedom' in place of 'free will'.

morally responsible for an action that it was not in his power to avoid. But if human behavior is entirely governed by causal laws, it is not clear how any action that is done could have ever been avoided.⁴⁷

Frankfurt's solution to the problem was to argue the PAP is false, and he has provided famous thought experiments showing why. His version of compatibilism is one of two major varieties I will present in the next section. The other comes from P.F. Strawson.

1.7 Compatibilism Introduced

Before we enter into the minutia of Frankfurt's and Strawson's accounts, I will present a brief history of compatibilism,⁴⁸ especially the accounts that set the stage for Frankfurt and Strawson. Recall, I will appeal to these and other compatibilist accounts, as a means of mitigating the negative social consequences predicted in the recent psychological research.

The following history of compatibilism is by no means comprehensive. I include a couple highlights that act as a backdrop for the compatibilism of Frankfurt and Strawson. I will delineate two versions of what is commonly referred to as *classical compatibilism*. The first comes from Hobbes and the latter is rooted in Hume but is often associated with A.J. Ayer. The reason why Frankfurt and

⁴⁷ Ayer (1954), p. 271.

⁴⁸ Sometimes used interchangeably with 'soft determinism', a term coined by William James in 1884. However, strictly speaking, 'compatibilism' (coined by Keith Lehrer) refers to the compatibility of determinism and free will. A compatibilist need not assert that either view is the case, while a soft determinist maintains that determinism *is* true and is compatible with free will.

Strawson's accounts have been at the forefront of contemporary literature is because they both serve as lifelines for the compatibilist in the wake of the failure of classical compatibilism. In fact, most who work in free will consider the objections to Hobbes' and Ayer's accounts to be insurmountable. Let's begin by looking at Hobbes' view.

1.8 Classical Compatibilism: Hobbes

In *Leviathan*, Hobbes argues a person has free will if there is “no stop, in doing what he has the will, desire, or inclination to do.” He is placing two conditions on free actions: (a) that a person acts on a desire or inclination and (b) this action is not ‘stopped’ or prevented. Given the first condition, a free action is one a person wills; given the second, a free action mustn't be impeded, e.g., by an external force. The following is a case that meets the first but not the second condition. Suppose after being the passenger of a long car ride, I have a desire to get out of the car. The car is parked, but the child safety lock, controlled by the driver, prevents me from exiting. I'm sitting in the car, but not freely so. According to Hobbes, this is because acting on my desire to leave is thwarted by the driver's handle on the safety lock. The person in the car isn't free, nor is anyone who isn't doing what she desires. Take, for instance, a prisoner cleaning toilets with a toothbrush. We ask the prisoner “are you doing what you desire?” “No,” he replies. The prisoner's act of cleaning can't be free on Hobbes' definition. Never are you free if you're not doing what you want. Such actions fail to satisfy the first condition.

Perhaps there are obvious counterexamples to Hobbes' analysis. For instance, there are many things I do, which I consider to be free, that I don't want to do. Driving to jury duty or taking out the trash, for instance. But since I don't want to do them, they can't be free on Hobbes' analysis. What someone defending Hobbes might say in reply is that while it's true I don't enjoy jury duty or taking out the trash, I still freely perform these acts if I desire them instrumentally. For example, I desire to drive to jury duty as a means to avoiding the penalty I would otherwise incur. I desire to take out the trash as a means to the fulfillment of another desire: the desire for a clean home.

Setting the issue aside, Hobbes' analysis is problematic for other reasons. His opponent argues it's too broad because there are actions that meet both Hobbes' conditions yet aren't free. Suppose while in the midst of grading philosophy exams, I get thirsty. I decide to take a sip from my water bottle. Does my taking a sip meet both of Hobbes' conditions? Surely, it does. I have a desire to take a sip of water and nothing impedes acting on that desire. But—on the assumption that determinism is true—was my taking the sip free? It depends on your intuitions about free will. Some, like Hobbes, have the intuition that actions like taking a sip of water are acts of free will. However, others disagree. There is a strong intuition, especially among incompatibilists—those who assert that determinism and free will are incompatible—that a necessary condition for free will is the ability to do otherwise than what one in fact did. On the assumption of determinism, I can't be free when I take a sip of water because I couldn't have done otherwise. The same goes for all actions that meet

Hobbes' conditions. Is this intuition about the ability to do otherwise condition justified? It is not my purpose here to say. For now, I will briefly mention that a sizeable portion of the literature on compatibilism is dedicated to the ability to do otherwise condition, and I will assume for the sake of argument that Hobbes' analysis is inadequate. As discussed, for Hobbes a free action is one that proceeds from the will and is not impeded upon by external factor. Hobbes himself didn't think that the ability to do otherwise condition was a requirement of free will; in fact he believed the notion to be nonsensical. If my taking the sip of water is determined in virtue of prior conditions and the laws of nature—or in Hobbes' language “when all things are present which are needful to produce the effect”—for me to be able to do otherwise “implies a contradiction and is nonsense.”⁴⁹

1.9 Conditional Compatibilism: Ayer

A second, more promising strand of classic compatibilism is seen in the work of G.E. Moore and A.J. Ayer but is rooted in Hume. All three claim individuals *do* have the ability to do otherwise—at least in a sense—even if determinism is true. Ayer, for instance, argues we have what has been described as a *conditional* ability to do otherwise. A person has a conditional ability to do otherwise if she performs some action but would have done otherwise counterfactually, i.e., if the world had been different. It's a conditional ability because it depends on a counterfactual condition.

⁴⁹ Hobbes (1654), *Of Liberty and Necessity*, §32.

Given the actual past, adequate determinism may necessitate our lives unfold in a particular, unavoidable way. However, *counterfactually*, i.e., if the past had been different, events could have unfolded differently. Given the actual past and the laws of nature, it may be unavoidable I will have a ham sandwich for lunch tomorrow. *But*, if the past had been different, I'm not locked into this lunch item; I could have done otherwise. Perhaps I would've had a PB&J instead. That example shows the conditional ability to do otherwise can exist in a deterministic world.

The conditional ability to do otherwise is but one piece of the puzzle for Ayer's analysis of free will:

To say that I could have acted otherwise is to say, first, that I should have acted otherwise if I had so chosen; secondly, that my action was voluntary in the sense in which the actions, say, of the kleptomaniac are not; and thirdly, that nobody compelled me to choose as I did: and these three conditions may very well be fulfilled. When they are fulfilled, I may be said to have acted freely.⁵⁰

He claims three conditions must be met to act freely. The first is the conditional ability to do otherwise condition. The second is that the act must be voluntary. Thirdly, the act mustn't be compelled. When he specifies the first condition, Ayer writes, "I should have acted otherwise if I had so chosen." Here's an example to explain what this means. Suppose, on Ayer's analysis, I freely ordered chocolate ice cream at an ice cream parlor. Meeting his ability to do otherwise condition means if I had chosen differently, if I had chosen vanilla ice cream, for example, I should have acted otherwise.⁵¹ But how could I have chosen the vanilla ice cream, if I am

⁵⁰ Ayer (1954), p 283.

⁵¹ Ayer uses 'should', but I take him to mean 'would'.

determined to order chocolate? If I'm determined to order chocolate ice cream, it would be impossible to order vanilla. *But* if I *had* ordered vanilla *and* determinism is true, then as far as Ayer is concerned the past *must have been different*. Whenever I act, it's *actually* true that I could have done otherwise if I had so chosen, i.e., if the past were different. When I order chocolate ice cream, Ayer can say I'm free because I do it voluntarily, I'm not compelled, *and* had the world been different I could have ordered something else. Put another way, the fact that I ordered chocolate ice cream in conjunction with the truth of adequate determinism does not negate the truth of the proposition *If I had chosen to order vanilla ice cream, I would have*.

The critics of Hobbes purported his analysis to be unsatisfactory because it didn't include an ability to do otherwise as a necessary condition for free will. Ayer provides an analysis of free will that *does* include this condition and goes further to show free acts on his analysis can occur in a determined world. However, Ayer's opponents argue his conditional ability to do otherwise is inadequate. For instance McKenna (2009) claims conditional analyses of the ability to do otherwise merely offer a 'hollow freedom'. McKenna concedes it's possible to possess a conditional ability to do otherwise even if the world is adequately determined, but questions whether this ability is the kind required for free will. Roderick Chisholm (1964) and many others have offered counterexamples to Ayer's account, i.e., examples of an action that is intuitively not free, but one that Ayer is forced into counting as free.

The expression (a) He could have done otherwise, it is argued, means no more nor less than (b) If he had chosen to do otherwise, then he would have done otherwise... The truth of statement (b)... is consistent with determinism...for even if all of the man's actions were

causally determined, the man could still be such that, *if*, he had chosen otherwise, then he would have done otherwise.⁵²

Chisholm argues that (a) and (b) are not equivalent. The ability to do otherwise that Chisholm claims is a requisite for free will is not satisfied by someone with Ayer's conditional ability to do otherwise. Moreover, Chisholm happily admits if one had chosen differently (had the past been different), it's true one would have done otherwise, but that in conjunction with Ayer's other two conditions for freedom is still insufficient.

Given determinism, prior conditions in conjunction with natural laws necessitate one's choices. If I chose chocolate ice cream at the parlor, that decision was necessitated by prior conditions in conjunction with the laws. Appealing to counterfactual circumstances will not change this necessitation. The kind of ability to do otherwise needed for freedom, Chisholm argues, is the kind that gets around the physical necessitation of determinism. A conditional ability to do otherwise fails in this regard. Thus, Chisholm maintains there is an important distinction between someone who could have done otherwise and someone who, if he had chosen to do otherwise, would have done otherwise. For Chisholm, if I freely choose chocolate ice cream, then I can choose chocolate or choose to do something else (like order vanilla). Both options are actually available, that is, either may obtain. Chisholm is an incompatibilist. For him free will and determinism are not compatible because to have the sort of ability to do otherwise required for freedom requires that some human actions are not determined.

⁵² Chisholm (1964), p 175.

Others, too, have questioned whether or not a conditional ability to do otherwise is the sort of ability that should be required for free will. Many including Chisholm and Lehrer claim that it isn't. Here is a variation of a counterexample to Ayer-like accounts from Lehrer (1968).⁵³ Suppose, Lehrer has us imagine, Sara is presented with a bowl of licorice, some of the pieces being black and some being red. Unbeknownst to Sara, she has a psychological pathology making her averse to eating red candy. Perhaps it reminds her too much of blood. Sara chooses the black licorice. Notice both of the following propositions are true:

- (1) *Sara could not have eaten a piece of red licorice.*
- (2) *If Sara had wanted to eat a piece of red licorice, she would have.*

What does this mean in terms of our competing analyses of ability to do otherwise? On a *traditional* analysis, Sara does *not* have said ability, but on the *conditional* analysis, Sara *does* have the ability to do otherwise. But this doesn't jibe with most people's intuitions, Lehrer maintains. Most people would say that in terms of her licorice choices, Sara chose the black licorice and could not have chosen the red, i.e., she couldn't have chosen otherwise. However, according to a proponent of the conditional analysis, Sara, could have. But does she have the ability to choose the red licorice, in an intuitive sense? Most people would say no, and Lehrer, Van Inwagen, and Chisholm, among others claim this is compelling evidence the conditional analysis of the ability to do otherwise should not be included in our analysis of free will.

⁵³ p. 32.

I've briefly sketched two classical compatibilists accounts, both of which face serious objections. The first was Hobbesian compatibilism, which defines a free action as the result of an individual's unstopped desire. Opponents of this view worry Hobbes' definition is too broad. That is, they believe there are actions Hobbes must count as free, which aren't intuitively free because the individuals responsible for them do not have the ability to do otherwise. Moore, Ayer, etc. argued that, we do, in fact, have the ability to do otherwise—if we think of that ability conditionally. I covered two objections to that sort of approach. One from Chisholm, who claims if 'ability to do otherwise' merely means 'I would have done otherwise, had I wanted to', then we are forced to call actions free that aren't. The second objection was a Lehrerian counterexample intended to show that Ayer's conditional ability to do otherwise should not be a part of the correct analysis of free will.

1.10 Contemporary Compatibilism: Frankfurt

A ray of hope for compatibilists emerged in the 1960s. Harry Frankfurt published a "remarkable and now classic"⁵⁴ paper entitled "Alternative Possibilities and Moral Responsibility," in which he famously claimed the PAP is false, that is, that determinism can be true and we can have free will *without the ability to do otherwise*. The paper included several counterexamples to the PAP, cases in which

⁵⁴ Van Inwagen (1999), p. 343.

an individual clearly did not have the ability to do otherwise but intuitively had free will. One of them goes like this:

Black is such that he wants Jones to perform some action—call it A. If Jones performs A on his own, Black will take no action. However, if it's clear that Jones will do something other than A, then Black will prevent him, and, further, will force Jones into doing A. How will Black force Jones into doing A? Frankfurt imagines this could be achieved by secretly slipping Jones a potion or through hypnotic suggestion, such that it would cause in Jones an irresistible urge to do A and nothing else. Black might even accomplish this by manipulating “the minute processes of Jones’s brain and nervous system in some more direct way, so that causal forces running in and out of his synapses and along the poor man’s nerves determine that he chooses to act and that he does act in the one way [A] and not in any other.”⁵⁵

Now what if Black, intently watching Jones, monitoring his brain processes, and waiting for the slightest indication that he will do *not* A, observes Jones performing A with no intervention on his part? The question, then, Frankfurt asks: Intuitively, did Jones freely do A? We will tend to answer, “yes,” Frankfurt argues. In my introduction to philosophy courses, I present undergraduates with Frankfurt-style thought experiments, and the vast majority tend to believe that Jones freely performed A. Frankfurt’s thought experiment seems to serve as a compelling counterexample to the PAP, that is, it captures a case in which an individual performs

⁵⁵ Frankfurt (1969), pp. 835-836.

an action freely, *despite not having the ability to do otherwise*. It's clear Jones can do A and nothing else. For before he can perform any other action, Black will intervene.

There are many variations on Frankfurt's thought experiment. Here is one from Van Inwagen:

Suppose that Gunnar has shot Ridley and is morally responsible for having done so. (Build into the example whatever you think is needed to make this supposition true.) Now add to the case an offstage "counterfactual manipulator," Cosser, who would have caused Gunnar to shoot Ridley (perhaps by direct manipulation of Gunnar's brain) if Gunnar had shown any hesitation about carrying out his long-standing plan to shoot Ridley. (In saying that Cosser is an "offstage" manipulator, we mean that his existence, powers, and intentions are unknown to Gunnar, and that, unless Cosser were forced to carry out his "contingency plan," nothing he did would have any effect on Gunnar). But, in the event, Gunnar showed no such hesitancy and, as we have said, went ahead and shot Ridley.... But then Gunnar is morally responsible for having shot Ridley even though he was unable not to shoot Ridley.⁵⁶

Van Inwagen in his conclusion uses the phrase 'morally responsible', but I think he could also conclude that Gunnar freely shot Ridley, even though he was unable not to (i.e., he couldn't have done otherwise). Gunnar's being morally responsible for having shot Ridley implies he freely shot him. I repeat what I said in an earlier footnote: there are many who claim that if someone is morally responsible for having done something, then she freely did it. Thus, Van Inwagen's version of Frankfurt's thought experiment leads to a similar conclusion: there are cases in which a person can have free will, despite lacking the ability to do otherwise. Frankfurt's paper inaugurated an era of contemporary compatibilism. For the first time in years,

⁵⁶ Van Inwagen (1999), p. 344.

compatibilists seemed to have the upper hand against the incompatibilist, whose argument might be best construed as the following:

- (1) If adequate determinism is true, then no one ever has the ability to do otherwise than she, in fact, does.
- (2) If no one ever has the ability to do otherwise than she, in fact, does, then no one has free will.
- ∴ (3) If adequate determinism is true, then no one has free will.

If (1) and (2) are true and it's true that there is strong evidence to support adequate determinism, it is likely, then, that no one has free will. Frankfurt's paper put a hole in the incompatibilist's sails by providing a counterexample to (2). Today, it is still an open debate whether or not he succeeded.⁵⁷

1.11 Frankfurt's Account of Free Will

In addition to combating the PAP, Frankfurt offered another seminal paper offering a *positive* account of free will. In "Freedom of the Will and the Concept of a Person," he outlines the criteria for a free action. His analysis begins with the concept of a *person*. Desires are to be found in all sorts of animals, many of whom would not suitably be called persons, Frankfurt notes, but 'second-order desires' are "peculiarly characteristic of humans." A first-order desire is a desire for anything (except for a desire). A second-order desire is a desire to have or not have a certain desire. When one wants a first-order desire to be effective, that is, to move her to

⁵⁷ I will not pursue this further, but for more on this see Naylor (1984), Ginet (1996), and Mele and Robb (1998).

action, Frankfurt calls this second-order desire a second-order *volition*, which is necessary for being a person. A *wanton* is someone who has first-order desires, (and perhaps second-order desires) but no second-order volitions. Examples of wantons for Frankfurt are nonhuman animals, very young children, and even some human adults. Frankfurt gives an example of an ‘unwilling addict’, someone who has a desire for a drug (i.e., a first-order desire), but wishes to not have this desire (i.e., a second-order desire). He imagines, then, the person also has a first-order desire *not* to take the drug. And although she desperately wants the latter first-order desire to win out and be the effective desire that she acts upon, she acts on the former desire and consumes the drug.

A wanton addict has no second-order desire about her desire for narcotics. She is moved by a first order desire for the drug and never considers whether or not she wants to be moved by such a desire. A lab mouse addicted to cocaine is a good example of a wanton addict. The mouse acts in ways necessary in order to secure consumption of the drug, but presumably never considers whether or not it *wants* to act on that desire.

For Frankfurt, a person has free will only if she has second-order volitions. To say that a person enjoys freedom of the will means “he is free to will what he wants to will, or to have the will he wants.... It is in securing the conformity of his will to his second-order volitions, then, that a person exercises freedom of the will.” That is, a person has free will only if she can align her first-order desires with her second-order volitions. We can now see that Frankfurt includes his talk of the

wanton and unwilling addict to contrast them with a person who exercises free will. The wanton addict doesn't have free will because she has no second-order volitions with respect to drugs. The unwilling addict does not have free will when she consumes the drug because the first-order desire that leads her to action does not align with her second-order desires. That is, she wants the drug and consumes it because of that strong desire; however, she wishes not to act upon it. So while she has a second-order desire, she does not possess a second-order *volition*. For her to have a second-order volition, in this case, her first-order desire to not take the drug would have to win out, which would in turn conform to her second-order desire, thus making it a second-order volition and marking her freedom.

1.12 Objections to Frankfurt

There are salient objections to Frankfurt's account of free will. One of the most prominent comes from Watson (1975). Watson objects to the special status Frankfurt places upon second-order volitions. What makes *second*-order volitions so special when it comes to free will? Why not give the status to third, fourth, fifth, etc.? Watson questions the special status of *second*-order volitions and claims that in some cases whether or not one has free will depends whether or not one can align her second-order desire with a *third*-order volition. Take the drug case, for instance. Suppose I have conflicting first-order desires: I want and don't want to take the drug. Couldn't I also possess conflicting *second*-order desires? Perhaps I want to want the

drug (I want to be a cool rock star who does drugs) and also don't want to want the drug (I don't want drugs to ruin my life). In addition to these conflicting sets of desires, I only want my second-order desire (that I don't want to desire drugs) to be effective. I succeed in aligning this *third*-order volition with this second-order desire. In that case, free will seems to depend on the alignment of this third-order volition. Moreover, we could imagine cases in which even higher order volitions need to be taken into consideration.

Frankfurt was aware of this problem and offers the following reply:

It is possible ... to terminate such a series of acts [i.e., ever higher-order volitions, e.g., third, fourth, fifth, etc.] without cutting it off arbitrarily. When a person identifies himself decisively with one of his first-order desires, this commitment "resounds" throughout the potentially endless array of higher orders... The fact that his second-order volition to be moved by this desire is a decisive one means that there is no room for questions concerning the pertinence of volitions of higher orders... The decisiveness of the commitment he has made means that he has decided that no further question about his second-order volition, at any higher order, remains to be asked.

In other words, when someone decisively identifies with a first-order desire, he finds no need to be concerned about volitions beyond the second order. Those higher-order volitions (if they exist) are not relevant, i.e., volitions of higher orders are of no concern, he decides.

Frankfurt uses an example to clarify this. He has us imagine a person who has a second-order volition, such that the person wants her desire to concentrate on work to be effective. The volition is 'decisive', meaning if we were to ask the individual, for instance, if she *wants* to want to want to concentrate on her work, she would respond by saying the question does not arise in her. In other words, she has decided

that no further investigation into high-order desires is necessary because it's a non-issue, such a desire would not be pertinent.⁵⁸

1.13 Contemporary Compatibilism: Strawson

Let's now turn to P.F. Strawson's compatibilism in "Freedom and Resentment." As we shall see, Strawson's approach to free will is quite different from Frankfurt's but has attracted just as much attention in the literature. Many claim Strawson doesn't offer a traditional analysis of 'free will' but rather a *descriptive* account of moral responsibility by way of human psychology. Russell (1995) and McKenna and Russell (2008) argue there is an echo of Hume in Strawson's writings. Both carefully consider the relationship between human nature (specifically, our emotions and psychological dispositions), moral responsibility, and free will. Early in the essay, Strawson focuses on what he considers common human responses. When we are wronged, we commonly experience a reaction of resentment, he says. When we see another wronged, we form an attitude of moral indignation or disapprobation. And when we have wronged others, we feel guilt. Strawson casts his account in terms of these basic sorts of reactions. Central are what he calls *reactive attitudes*. Reactive attitudes are "essentially natural human reactions to the good or ill will or indifference of others towards us, as displayed in their attitudes and actions."⁵⁹

⁵⁸ I will not pursue this further here. For further objections to Frankfurt's account see Slote (1980), Widerker (1995), Rostbøll (2010), and McKenna (2011).

⁵⁹ Strawson (1962), p. 195.

Elsewhere he describes them as “a complicated web of attitudes and feelings which form our moral life.”⁶⁰ Strawson says we form reactive attitudes in our interpersonal relationships. Attitudes like “resentment, indignation, hurt feelings, anger, gratitude, reciprocal love, and forgiveness,” without which we would hardly be human.⁶¹

Strawson calls some attitudes, *participant* reactive attitudes. They arise when one participant in a relationship sees another participant as a candidate for moral responsibility. However, if an individual held responsible is incapable of participating in such a relationship—someone with an abnormal or underdeveloped psychology, a mentally ill or a young child, for instance—then our attitudes take on what Strawson calls an *objective* stance. Understanding objective attitudes is key to comprehending how Strawson deals with the problem of determinism.

Strawson says we must ask ourselves, “What effect would, or should, the acceptance of the truth of a general thesis of determinism have upon these reactive attitudes?”⁶² Would or should determinism result in “decay or the repudiation” of these attitudes like “gratitude, resentment, and forgiveness.”⁶³ Strawson thinks not. That consequence is most likely psychologically impossible because our commitment to them and the relationships in which they form is “too thoroughgoing and deeply rooted.”⁶⁴

Strawson shows why he thinks we are justified in keeping personal reactive attitudes, even if determinism were true, by using several examples. One is about a

⁶⁰ Strawson (1962), p. 205.

⁶¹ Eshleman (2009).

⁶² Strawson (1962), p. 195.

⁶³ *Ibid.*

⁶⁴ *Ibid.*, p. 196.

psychoanalyst and his mentally ill patient. The psychoanalyst takes on what Strawson calls an 'objective stance' toward the patient. Taking an objective stance is to treat someone as we would a person who cannot engage in ordinary interpersonal relationships. The point, of course, of their sessions is to cure the patient of this illness, thereby making him a normal participant in society who could engage in ordinary personal relationships—which entail being the object of personal reactive attitudes. But if it's true that determinism threatens our free will and negates moral responsibility, then, as Strawson points out, the entire enterprise of psychoanalysis (or any sort of psychological counseling) is pointless—at least in terms of helping patients to become responsible for their actions. If moral responsibility is nonexistent, then a patient's behavior is excused (that is, she could not be held responsible for the behavior) before the treatment, *as well as after*. As I see it, Strawson's argument is as follows:

(1) The enterprise of counseling psychology/psychiatry is often successfully in helping patients become autonomous, independent, socially competent, functioning members of society, such that the patient would be held responsible for her subsequent actions (i.e., she would be subject to personal reactive attitudes).

(2) If the truth of determinism is a genuine threat to free will and responsibility, then the enterprise of counseling psychology/psychiatry has never been successful in helping patients to become autonomous, independent, socially competent, functioning members of society, such that the patient would be held responsible for her subsequent actions (i.e., she would be subject to personal reactive attitudes).

∴ (3) The truth of determinism is *not* a genuine threat to free will or responsibility.⁶⁵

⁶⁵ Strawson (1962), pp. 195-195. Some might argue Strawson begs the question by assuming we are sometimes morally responsible for our actions.

Even if it were psychologically possible for us to take a completely objective standpoint in our reactive attitudes, Strawson thinks it would be irrational to give up personal reactive attitudes, for they carry an immense practical value. Giving them up would substantially lower the quality of our lives. As McKenna (2009) puts it, “the richness brought to human life by seeing persons as members of an interconnected community, and hence as morally responsible for their conduct, would far outweigh any other benefits that could be gained by giving up these practices.” In other words, if we gave up this sense of interconnected community on the acceptance of determinism, life would hardly be worth living.

1.14 The Metaphysics of Strawson’s Account

I have discussed Strawson’s overall project and what he sees as the relationship between reactive attitudes, responsibility, and the threat of determinism. I will go into the metaphysical issues that arise vis-à-vis these attitudes. Strawson holds reactive attitudes do not track some sort of objective responsibility independent of our attitudes and practices (there is no such thing on his view). Our attitudes alone, instead, *determine* responsibility. To explain what I mean by what I’ve said in the last two sentences, I will appeal to the term ‘truth-maker’ introduced by Mulligan, Simmons, and Smith (1984), the referent of which is at the heart of Russell’s

correspondence theory of truth.⁶⁶ A truth-maker is that entity “in virtue of which sentences and/or propositions are true.”⁶⁷ Examples of truth-makers are facts and states of affairs. Sentences and or propositions are examples of *truth-bearers*. The truth-maker of the proposition, *The Sun has a greater mass than the Earth*, is *the fact* that the Sun has a greater mass than the Earth. Notice in this case the truth-maker is independent of human attitudes and practices. Moreover, *the Sun has a greater mass than the Earth* is true *irrespective* of our attitudes and practices.

Some truth-makers *are* dependent upon such attitudes and practices. For example, the truth-maker of *The audience found the comedian hilarious* depends on the attitudes of the audience. Take, however, the sentence, “Kim is morally responsible for shooting Miguel.” There are many moral philosophers—some moral realists for instance—who would claim this sentence expresses the proposition, *Kim is morally responsible for shooting Miguel*, the truth-maker of which is independent of our attitudes and practices. For them, *Kim is morally responsible for shooting Miguel* is true in virtue of the fact that Kim is morally responsible for the shooting. If we imagine that—for whatever reason—there were no reactive attitudes (other things being equal), these moral realists could still maintain that Kim is morally responsible for shooting Miguel. Or suppose reactive attitudes exist, but they were such that the individuals who possessed them did *not* hold Kim morally responsible. We could imagine this being the case in a courtroom in which neither the jury, judge, nor

⁶⁶ We do not need to assume Russell’s theory is correct; I am merely appealing to the idea of truth-makers to make sense of Strawson.

⁶⁷ Mulligan, Simmons, and Smith (1984), p. 287.

members of the community found Kim guilty. The realist can argue these attitudes don't determine the facts and, hence, do not have any bearing on *the fact* (as we posit for this case) that Kim is responsible for shooting Miguel. Strawson, on the other hand, does not hold this sort of realist view and instead thinks in general "what it is to hold a person morally responsible for wrong conduct *is* nothing more than the propensity towards, or the sustaining of, a morally reactive attitude of disapprobation." And this attitude of disapprobation is a response to an observed ill will from the individual held responsible.⁶⁸

As opposed to the aforementioned sort of realism, Strawson does not hold that facts independent of our attitudes are the truth-makers of propositions about responsibility. Instead, responsibility consists in reactive attitudes expressed by those who hold others responsible. Hence, in the case of Kim and Miguel, there is no fact about moral responsibility independent of reactive attitudes. "Only by attending to this range of attitudes," Strawson asserts, "can we recover... a sense of what we mean, i.e. of all we mean, when, speaking the language of morals, we speak of desert, responsibility, guilt, condemnation, and justice," and we "do not have to go beyond" attending to this range. This last bit is important: for we learn the buck stops at the attitudes, that is, we needn't seek any facts independent of our attitudes to ground our concepts of responsibility, desert, freedom, etc. because on Strawson's view *there are none*. As Eshleman writes, "*Judgments about being responsible* are understood in relation to the role reactive attitudes play in the *practice of holding responsible*, rather

⁶⁸ McKenna (2009).

than the other way around [author's emphasis]."⁶⁹ This is a reversal of the traditional view. On the traditional view, when we form an attitude about whether or not someone is responsible, the appropriateness of that attitude is based upon whether it would be correct to judge, in an objective sense, the person as responsible. But as Strawson sees it, it's the other way around. Whether or not it would be 'correct' to judge a person responsible depends solely on our attitudes about responsibility.

How did the traditional view come to be? Strawson maintains there has been "over-intellectualization" of the concepts like desert, responsibility, guilt, etc. This over-intellectualization is to blame for the fact that philosophers think, for instance, that people are responsible only when they have "satisfied some set of objective requirements on being responsible."⁷⁰ But there is no set of objective requirements for responsibility or any of the related concepts, for Strawson. These objectivist thinkers also hold that we are *justified* in holding individuals responsible only if they are actually responsible—independent of our attitudes. However, Strawson asserts the practice of holding persons responsible, which is so deeply a part of our lives, "neither calls for nor permits, an external 'rational' justification."⁷¹ The "external 'rational' justification" to which Strawson is referring could be an attitude-independent fact or the satisfaction of some objective criteria. Hence, in the Kim/Miguel example, what justifies us in holding Kim responsible isn't that our attitudes track some objective fact about responsibility (again, there is no such thing

⁶⁹ Eshleman (2009).

⁷⁰ *Ibid.*

⁷¹ Strawson (1962), p. 205.

on Strawson's view). If we hold Kim responsible, then he *is*; no further justification is necessary. There are many details to be filled in, and Strawson himself acknowledges his paper serves as more of a starting point than a precise and comprehensive conceptual analysis.

1.15 Problems for Strawson

There are several pressing objections to Strawson's account; however, I will only tackle one. But first I want to note, as Eshleman does, that while Strawson's view is a form of compatibilism, it doesn't seem to *resolve* the problem of free will, that is, it does not show that objective conditions for free will and responsibility can be satisfied, given the truth of determinism. Instead, Strawson *dissolves* the problem of free will by arguing the social practice of holding persons morally responsible requires no objective conditions and "therefore needs no external justification in the face of determinism."⁷²

Fischer and Ravizza (1993) argue, as Eshleman puts it, if "being responsible is defined in relation to the practice of holding responsible," then it would be impossible to criticize practices of holding persons responsible from a standpoint *outside* of them.⁷³ Suppose an individual argued for the modification of a particular practice, apartheid in South Africa, for instance. Suppose this individual were alone in this view. Because being responsible goes no deeper than attitudes about responsibility,

⁷² Eshleman (2009).

⁷³ *Ibid.*

there is nothing external from these attitudes for the individual to support her criticism. If the person had such grounds, this would entail some sort of external constraint on responsibility beyond the reactive attitudes. However, this would, in effect, vitiate Strawson's theory because he maintains there are no "independent theoretical conditions on being responsible."⁷⁴

The problem as Fischer and Ravizza see it is:

...Strawson's theory may reasonably be said to give an account of what it is for agents to be held responsible, but there seems to be a difference between being *held* responsible and actually *being* responsible. Surely it is possible that one can be held responsible even though one in fact is not responsible, and conversely that one can be responsible even though one is actually not treated as a responsible agent. By understanding responsibility primarily in terms of our actual practices of adopting or not adopting certain attitudes towards agents Strawson's theory risks blurring the difference between these two issues.⁷⁵

A reply to this from McKenna and Russell (2008) goes as follows: Strawson does *not* suggest that someone is responsible *whenever* she is held responsible. That is, "in particular cases our reactive attitudes may fail to properly track moral responsibility."⁷⁶ For instance, maybe we misunderstand a person's intentions, or maybe our attitudes are inappropriate because we don't have all the facts.

However while Fischer and Ravizza may accept Strawson's reply about these *particular* inappropriate reactive attitudes, they don't see how it could work if there existed a *systematic* lack of connection between the attitude holders and the object of their attitudes, "for example, that an entire community has its reactive attitudes

⁷⁴ Eshleman (2009).

⁷⁵ Fischer and Ravizza (1993), p. 12.

⁷⁶ *Ibid.*

switched on or off in the wrong way and at the wrong times.”⁷⁷ We could imagine this happening, say, if a malicious being tinkered with our DNA such that we all became ‘social monsters’, or if we were subject to intense propaganda and social conditioning like that in Nazi Germany. If there were no constraint beyond our attitudes and practices, then a Strawsonian may be committed to the view that criticizing such ‘sociopathic’ communities is nonsensical. I set this issue aside for now and will return to it in the final chapter.

1.16 Contemporary Compatibilism: Other Accounts

I have discussed two accounts of contemporary compatibilism, namely the accounts of Frankfurt and Strawson. This is not to suggest there are only two, or that these are the only two that attract much attention in the literature. There are several other promising compatibilist accounts, which I have not mentioned. I will briefly mention them now and say a couple of things about each without going into much detail. One view responds to the criticism made against the conditional accounts of Moore and Ayer. For instance, Kadri Vihvelin argues that if one appeals to a more *nuanced* account of dispositions—as opposed to the simple counterfactual conditions that we see in Moore and Ayer—we can salvage the account. Her account is sometimes called *the new dispositionalism*.⁷⁸ Recall, the problem with conditional

⁷⁷ McKenna (2009).

⁷⁸ For more on this see, Vihvelin (2004), Fara (2008), and Clarke (2009).

accounts, like those from Moore and Ayer, is that they are subject to serious counterexamples, like the one from Lehrer about Sarah and the bowl of licorice. Sarah is unaware of her psychological aversion to red licorice, and when she is offered a choice between red and black licorice, she chooses black. Lehrer says most people's intuition is that Sarah's selection isn't free because of her psychological aversion to red licorice. But the proponent of the conditional account is committed to the view that Sarah *did* freely choose the black licorice. The view is wrong, Lehrer purports, because it commits us to something we intuitively think is wrong, namely that Sarah freely chose the black licorice. The conditionalist claims Sarah acts freely in this case because it's true that if Sarah had chosen the red licorice, she would have done otherwise. But if Sarah had chosen the red licorice, then she wouldn't be psychologically averse to red licorice, which is to have us imagine a world that is not relevant to whether or not Sarah is free in *this* world. We want to know, *given* Sarah's pathology, if she freely chooses black. Simply forgoing that assumption by noting there is a world in which Sarah doesn't have said pathology and does choose red licorice is irrelevant with respect to her present licorice dilemma.

How does the new dispositionalist improve upon this account? According to Vihvelin (2004), to have free will is to have "the ability to make choices on the basis of reasons and to have this ability is to have a bundle of dispositions."⁷⁹ There are two claims here. The first, that free will is the ability to make a choice on the basis of reasons, Vihvelin takes to be uncontroversial. Let's grant her that and set it aside.

⁷⁹ p. 429.

Let's unpack her second claim, that to have free will is to have a bundle of dispositions. Vihvelin maintains that for a person to have a disposition, there must be *intrinsic properties* that are the *causal basis* of the disposition. What does this mean for an ordinary instance of free will, such as the licorice case, but in which Sarah has no aversion to either color licorice? Suppose in this case Sarah, unimpeded by any pathology, freely chooses red licorice. On Vihvelin's view, this means Sarah's intrinsic properties (her physiology, psychology, etc.) cause in her a disposition to choose red licorice and are such that she does choose red licorice in an "appropriately rich range of such counterfactual conditions" in which the case base operates unimpaired.⁸⁰ (How many counterfactual conditions comprise an appropriately rich range, let's set aside). Put another way, for an act to be free we consider a multitude of counterfactual worlds in which Sarah is caused by her intrinsic properties to have a disposition to choose red licorice and she does choose it in some of them. This version of conditionalism gets around the Lehrer objection because although Sarah chose the black licorice in the actual world, there are no counterfactual worlds where she is caused by her intrinsic properties (which includes those properties which cause her to have an unconscious aversion to red licorice), to have a disposition to choose red licorice.

Opponents to Vihvelin's account have questioned whether it is appropriate to restrict our consideration of counterfactual worlds only to ones in which one's dispositions operate unimpaired. McKenna (2009) notes that doing this in Frankfurt-

⁸⁰ McKenna (2009).

style cases is problematic. Take the Frankfurt case in which Jones shoots Smith without the intervention of Black, but had Jones not chose to shoot Smith, Black would have intervened and forced him to. McKenna claims our intuition is that Jones shot Smith and couldn't have done otherwise. But if Vihvelin is right, then Jones *could* have done otherwise, because when we consider counterfactual worlds where Jones' dispositions operate unimpaired, there are worlds in which Jones doesn't shoot Black.

Daniel Dennett offers an interesting account of compatibilism on which because of the impossibility, as he sees it, of viewing the world purely in physical (deterministic) terms, we are justified in keeping our everyday, folk psychology notions of free will. Although human action is physically determined, he argues this is consistent with our taking what he calls a *personal stance* toward someone. Taking a personal stance toward someone is to make a pragmatic decision to treat her as a morally responsible person. Moreover, someone is morally responsible if taking the personal stance toward her "pays off." Dennett contrasts taking a personal stance toward someone with taking a physical or deterministic stance. To take the physical stance would be to view the individual only in basic physical terms (e.g., as a set of subatomic particles in motion). Dennett argues it is not practically possible to view individuals from the physical stance. In our everyday dealings with other people, the personal stance offers much more utility, he thinks.

With respect to the PAP, Dennett (1984) agrees with Frankfurt: being able to do otherwise is not required for moral responsibility (or free will). In our everyday

dealings with other people, Dennett does not think considering whether or not one has the ability to do otherwise tells us anything interesting. He gives the historic proclamation of Luther as an example. When Luther, in front of the church door, claimed he “could do no other,” he didn’t mean he could not have done otherwise in the sense that incompatibilist mean, Dennett argues. Instead he likely meant something like he was morally resolute in his decision to protest.

For Dennett, free will consists in a person’s responsiveness to reasons, reasons that result in the evaluation and adjustment of their behavior. Beings with this capacity to respond to reasons arose naturally via a process of evolution, Dennett claims. In *Elbow Room*, Dennett argues for his variety of compatibilism by attacking the incompatibilist’s argument that no one is free because no one is the ultimate source of her actions. He objects to the imagery that is often employed to convey that premise, which often depicts humans as beings like marionettes or cogs in a machine. This imagery or ‘intuition pump’ is but distorting rhetoric, says Dennett, that serves to trigger faulty intuitions.⁸¹

The worry about our not ultimately being the source of our ‘free’ actions is not entirely groundless and has attracted much support in the literature. The response from incompatibilists is that Dennett’s dismissal is unfair and too widely sweeping. Nevertheless, Dennett’s account is promising and demands attention. If he can carefully articulate his objection to the source incompatibilist, his account would be a

⁸¹ Mckenna (2009).

force to be reckoned with.⁸²

One of the most promising recent compatibilist approaches comes from John Martin Fischer and Mark Ravizza. They, along with many others who specialize in free will (including Dennett), put forth a *reason-responsiveness theory*, under which free actions are required to stem from rational consideration. Contrast a normal, healthy human with an individual who suffers from a psychological disorder impinging upon her ability to act from rational considerations. Take someone who engages in some sort of obsessive-compulsive behavior. When this individual person acts from her disorder, e.g., repetitively locking and unlocking doors, she is not acting freely on this view because she is not responding to rational considerations. Someone who acts freely, however, *is* sensitive to rational considerations. They don't, for instance, repetitively unlock and lock doors if there isn't reason to. But the view is more than being sensitive to rational considerations. According to Fischer and Ravizza (1998), a free action requires responsiveness to reasons in some *counterfactual* situations. For instance, if I freely make a peanut butter and jelly sandwich for lunch, then it must be the case that *if* in some counterfactual worlds I had a reason *not* to make the PB&J, I wouldn't have, e.g., if an old friend called me earlier in the day and invited me to eat out. On this view, the fact that I would have rationally considered the invitation and accepted is evidence that I am free when I

⁸² For more on Dennett's compatibilism, see Dennett (1973) and (1984). For critiques of his account, see Pereboom (2001), Mele (2006), and Ross (2011).

make the PB&J in actuality.⁸³

In the first section of this chapter, I presented major varieties of determinism: *causal*, *theological*, *logical*, and *bio-environmental*. The only kind of those four that I'm concerned with is causal determinism. We know that a determined world isn't necessarily a fully predictable world. We also know that causal determinism in a universal sense is most likely false on account of findings in quantum mechanics. However, because of *quantum decoherence*, the indeterminacy of the microscopic world or level has little to no bearing on the determinism of the macroscopic world. This kind of determinism we named 'virtual macro determinism' or 'adequate determinism'. Finally, I explained that there are some physicists and philosophers who think that adequate determinism is false with respect to human actions because of the chaotic nature of the human brain.

In the second section of this chapter, I presented many accounts of compatibilism. I began with compatibilist views from the history of philosophy: one from Hobbes followed by the conditional accounts of Ayer and Moore. Then I addressed the contemporary accounts of Frankfurt and Strawson: Frankfurt who famously claimed that determinism can be true and we can have free will despite lacking the ability to do otherwise and Strawson who holds free will and responsibility consist in personal reactive attitudes, which would either be psychologically impossible or irrational to give up on the acceptance of determinism.

⁸³ For more on Fischer and Ravizza's compatibilism, see Fischer (1987), Fischer (1994), and Fischer and Mark Ravizza (1998). For objections to their account, see Stump (1996), Mele (2000), and McKenna (2000).

I surveyed objections to all these accounts. There are many objections I did not cover, for to do so would be a project much different than the one in this chapter. There also exist many other varieties of contemporary compatibilism, which I either did not discuss at all or which I discussed only in brief. My purpose here was not to describe all these accounts in detail, nor to offer comprehensive coverage of every viable account of compatibilism.⁸⁴ Instead, this chapter served to (a) provide background on the important concepts addressed in the next chapter and (b) highlight evidence that there is at least one reason to think some human actions aren't determined and there is reason to believe in some form of compatibilism. As we will see in the next chapter, (b) may be important for mitigating the negative social consequences predicted by the work in psychology.

⁸⁴ Another recent compatibilist approach comes from Wolf (1990). Wolf offers a neo-Platonist compatibilist account on which one has free will when she acts in accord with "the True and the Good." Acting in such a way means acting in accord with reason. She calls this 'the reason view' on which if one is psychologically determined to perform good actions, she performs them freely and is deserving of praise. However, surprisingly, "being psychologically determined to perform bad actions is *not* compatible with deserving blame [my emphasis]."

2. Free Will and Alief

Of the varieties of determinism previously discussed, I will concern myself only with *scientific* determinism in this chapter. Following Van Inwagen (1975), scientific determinism can be expressed as the conjunction of the following:

(T1) For every instant in time, there is a proposition that expresses the state of the world at that instant.

(T2) If *A* and *B* are any propositions that express the state of the world at some instants, then the conjunction of *A* with the laws of physics entails *B*.⁸⁵

Scientific determinism—at least without qualification—is unlikely to be true.

Quantum physicists tell us on a small enough level, the world is not deterministic.

The behavior of electrons, for instance, can be indeterminate. However, because of quantum decoherence,⁸⁶ indeterminacy can be disregarded for the vast majority of macroscopic events, including the behavior of macromolecules, cells, organisms, rocks, tables, and humans. Behavior on this level is said to be ‘adequately’ or ‘virtually’ determined. Adequate determinism’s being true would explain how predictions in quantum mechanics virtually match those from classical physics with respect to the behavior of objects like billiard balls, the orbit of the Earth, and rocket ships propelled to the moon. That’s what determinism is.⁸⁷ The primary concern of this chapter is to address problems arising from a *belief* in determinism.

⁸⁵ p. 186.

⁸⁶ Quantum decoherence is the effect of a quantum system’s interaction with its environment. The details of decoherence get rather technical, and it is beyond the scope of this dissertation to explicate. For more on this, see Tegmark (2000) and Balaguer (2009).

⁸⁷ Here and henceforth I’ll use ‘determinism’ as a shorthand for ‘adequate scientific determinism’.

2.1 An Increase in Deterministic Thinking

Firstly, who does believe? Many in the scientific community do. Students in high school and college may, since they are exposed to determinism in their physics and chemistry courses. (At least determinism is an implicit assumption of those courses.) People on the street have perhaps a vague understanding of determinism. I suspect few people give it much thought—at least not on a day-to-day basis. Perhaps this is changing.

Scientific thinking has irreversibly woven itself into the tapestry of global culture. People trust science—or rather—the findings of scientists, and you don’t have to look farther than your pockets to see its impact. From tech devices to medicine, space exploration to the study of climate change, the world increasingly counts on and looks to science. That scientific thinking has become pervasive may have contributed to a significant increase in the number of people with a strong belief in determinism. Articles about determinism are appearing in the mainstream media with titles like, “Neuroscience, Free Will and Determinism: ‘I’m Just a Machine’”⁸⁸ and “Our Social Networks, Ourselves: Does Free Will Even Exist?: Scientists Are Finding that We’re Much More Predictable Than We Think.”⁸⁹ The first of those articles elucidates landmark work in neuroscience from Benjamin Libet, whose findings underscore the threat of determinism. Libet (1983) offers evidence that the

⁸⁸ Chivers (2010).

⁸⁹ Lehrer (2011).

non-conscious brain activity or *readiness potential* associated with action (like moving one's arm) occurs approximately half a second before the *conscious* will to act. Readiness potential is the neural activity in the motor cortex leading up to voluntary muscle motion. These findings alone don't guarantee determinism is true, but they don't bode well for proponents of free will. If awareness of these and similar studies is growing, evidenced by their appearance in mainstream media, belief in determinism may grow as well.

According to Rotter (1954), a person with a strong belief in determinism tends to attribute their behavior to external factors, such as environment and chance, as opposed to internal factors, like skill and effort. They perceive their 'locus of control' externally instead of internally. Twenge et al. (2004) found perceptions of locus of control have changed from the 1960s to 1990s. According to these researchers, the report of a perceived external locus of control from the 60s to 90s increased more than three quarters of a standard deviation. This may be the symptom of a growing belief in determinism.

If courts are supposed to represent the values and beliefs of the public, then another reason to think belief in determinism is spreading is that 'deterministic thinking' has increasingly found its way into the criminal law, according to Cotton (2005).⁹⁰ Cotton states, "By the [twentieth] century's third quarter, the criminal law had accepted a significant amount of deterministic thinking in virtually every one of

⁹⁰ In U.S. law, at least.

the areas in which the issue had arisen.”⁹¹ Related areas include: juvenile justice, expert witness testimony concerning mental state, the insanity defense, and sentence mitigation.

I’ve said what determinism is and provided reasons why belief in it could soon be widespread. I will now discuss the consequences of a strong belief in determinism. I begin with what I take to be general consequences followed by a discussion of consequences I consider specific. The first general consequence of a belief in determinism (sometimes I will omit ‘strong’) is that it sometimes undermines one’s belief in free will—especially if one is an incompatibilist. As a result, one’s motivation and sense of agency can be severely undermined.⁹² This can lead to a ‘why-bother’ mentality. Why bother putting in effort, developing skills, if behavior is ultimately governed by factors outside one’s control?⁹³ Further, a common belief is that free will is a necessary condition for morality. If there’s no free will, why bother acting ‘morally’?

Losing belief in free will may have other general consequences, too, as predicted by Paolo Verme (2009). The loss for Verme is couched in terms of Rotter (1954)’s *locus of control*. As noted, the locus depends on what Verme considers ‘external’ factors, such as fate, destiny, determinism, etc. (i.e., things ‘outside’ one’s control) or ‘internal’ factors like skill and effort. Verme found a weakened belief in free will causes a shift from an internal locus to an external one. A subject’s

⁹¹ p. 3.

⁹² This is suggested in Rigoni (2011).

⁹³ As evidenced by Baumeister and Vohs (2007).

perception of an internal locus of control was a better predictor of life satisfaction than “*any* other known factor such as health, employment, income, marriage, or religion.”⁹⁴ Verme’s findings are based on data from the European and World Value Surveys from 1981 to 2004 from eighty-four countries.⁹⁵

2.2 Determinism and Recent Work in Psychology

I discussed what I consider the more *general* byproducts of a strong belief in determinism. Recent work in psychology has indicated *specific* behavioral problems resulting from either a belief in determinism or a disbelief in free will.⁹⁶ For instance, Vohs & Schooler (2008) suggest a weakened belief in free will leads to increased cheating. In a follow up experiment, Vohs & Baumeister (2009) provide evidence that experimentally inducing a disbelief in free will leads to an increase in “cheating, stealing, aggression, and reduced helping.”⁹⁷ In the 2008 experiment, subjects were split into groups, in which members of one read a booklet of fifteen statements like, “A belief in free will contradicts the known fact that the universe is governed by lawful principles of science” and “Ultimately, we are biological computers—designed by evolution, built through genetics, and programmed by the environment.” Another set of participants read neutral statements such as, “Sugar cane and sugar beets are grown in 112 countries.” Afterward, participants in both groups were asked to

⁹⁴ Verme (2009), p. 1.

⁹⁵ Available at: <http://www.worldvaluessurvey.org>

⁹⁶ In support of my earlier point, one could count the fact that so many of these studies are cropping up as further evidence that belief in determinism is spreading.

⁹⁷ p. 231.

complete fifteen challenging—yet solvable—logic, reading comprehension, and mathematical problems from Graduate Record Examination practice exams and told they would each receive \$1 for every correct answer. The experimenter then made up an excuse to leave and told subjects they were to grade their own problems, pay themselves \$1 for every correct answer, and shred the answer sheets. The results: subjects who read deterministic statements were more likely to cheat (by paying themselves for wrong answers) than participants who read neutral statements.

Stillman and Baumeister (2009) associate an induced disbelief in free will with reduced learning—specifically the kind of learning occurring from emotions like guilt. As in the Vohs and Schooler (2008) experiment, subjects read a booklet of deterministic statements to induce a disbelief in free will or neutral statements, which would have no such effect. Then participants of both groups were asked to describe in detail an experience in which they hurt someone and rate the intensity of the guilt associated with the event on a scale of 1 to 7.⁹⁸ Finally, subjects were asked, “Do you feel you learned anything from this event?” also on a scale of 1 to 7. Stillman and Baumeister found subjects who read the deterministic statements were less likely to report learning from guilt. The researchers hypothesized a decrease in this kind of learning could increase unethical behavior.

Stillman et al. (2010) find a disbelief in free will detracts from workplace success. That was the conclusion of the first of two studies. The second shows those with a stronger belief in free will received better job evaluations from their

⁹⁸ Baumeister et al. (1994, 1995) used this prompt to illicit feelings of guilt.

supervisors. So there's reason to believe thinking you're free really does make you a better worker, and further, that a weak belief in free will leads to weaker job performance than those who do believe.

There's more. Findings from Rigoni et al. (2011), suggest inducing a disbelief in free will also reduces readiness potential associated with basic motor preparation, like the intentional effort behind moving a limb. These researchers were surprised to find an influence of 'high-level beliefs' (e.g., a disbelief in free will) on basic motor preparation (e.g., the effort that goes into moving a limb). Roughly speaking, these findings suggest a disbelief in free will can undermine effort.

Philosophers, too, maintain widespread disbelief in free will could have unfortunate social consequences. C.P. Snow claims "we have to believe in free will to get along."⁹⁹ Saul Smilansky, in "Free Will: From Nature to Illusion," agrees with Snow that the social ramifications of weakened free will beliefs could be dire.¹⁰⁰ And as Nadelhoffer and Matveeva (2009) put it, a disbelief in free will could mean "wide-reaching negative intrapersonal and interpersonal consequences."¹⁰¹ Few would find these findings and theories surprising. There is a deep-seated intuition that we have free will. In the words of Jerry Fodor: "If it isn't literally true that my wanting is causally responsible for my reaching [...] and my believing is causally responsible for my saying [...] then practically everything I believe about anything is false and it's the end of the world."¹⁰² This is hyperbole, but poignant, nevertheless.¹⁰³

⁹⁹ Snow (1983).

¹⁰⁰ Smilansky (2001).

¹⁰¹ p. 495.

¹⁰² Fodor (1990), p. 156.

The primary focus of this chapter is to offer a way to prevent or mitigate these negative consequences—general and specific. I will present several ways to do this. It may be that in the end none of them are fruitful. That would not mean, however, this chapter would be in vain. There would be value in *ruling out* the philosophical avenues presented here. And even if not ruled out, these avenues may face serious objections. Such objections and replies thereof will be carefully considered. If, in the end, serious worries remain, it wouldn't be for naught, however; others interested in the topic could exploit the potential of the proposal provided here. Before introducing potential solutions, it would be prudent to clearly motivate the problem. To that end, I will carefully explain why some of the studies I mentioned in brief are really things philosophers should be worried about, that is, that there is good evidence a belief in determinism (or disbelief in free will) *causes* things like reduced helping, increased aggression, and reduced learning from emotions like guilt. And further, that there is evidence the increase is substantial and widespread enough to impact social functionality. The rest of the chapter, then, will focus on how to mitigate the negative consequences predicted by this research.

2.3 Baumeister et al. (2009): Belief in Determinism and Increased Aggression and Reduced Helping

¹⁰³ Although, there are some 'hard determinists' who are comfortable in their belief in determinism and the nonexistence of free will. Derk Pereboom (1995), for instance, holds this view and maintains that belief in determinism needn't undermine our interpersonal relationship nor "subvert the commitment to doing what is right."

Results from Baumeister et al. (2009) suggest a strong belief in determinism causes reduced helping and increased aggression. A weak belief in free will has the same effect, the results indicate. In the first of three experiments, subjects were randomly assigned to a 'determinism', 'free will', or 'neutral' condition. The determinism group read a set of fifteen statements, such as "Science has demonstrated that free will is an illusion"; "All behavior is determined by brain activity, which in turn is determined by a combination of environmental and genetic factors;" and "All behavior is determined by brain activity, which in turn is determined by a combination of environmental and genetic factors." In the free will condition, subjects read statements including, "I demonstrate my free will every day when I make decisions" and "I have feelings of regret when I make bad decisions because I know that ultimately I am responsible for my actions." In the neutral condition subjects read statements like, "Pocket calculators became common items only after 1970," "Oceans cover 71% of the earth's surface," and "Alkaline power cells generally work longer than ordinary batteries." Each subject read fifteen of their condition's statements for fifteen minutes at a rate of one per minute.

After reading the set of statements, participants read six hypothetical scenarios depicting individuals in need of help (e.g., a homeless person in need of money, a fellow student in need of a cell phone, etc.). Subsequently, subjects were asked to indicate—in that moment—how likely they would be willing to help in each situation on a scale from 1 (not at all likely) to 9 (very likely).

The average participant in the determinism condition reported a help score of 5.33, with a standard deviation of 1.52. In the neutral condition the average was 6.23 ($SD = 1.28$), and the average score in the free will condition was 6.27 ($SD = 1.19$).¹⁰⁴ Consider the difference between the determinism and neutral conditions, in which the average helpfulness dropped from 69.2% to 59.2% (6.23 to 5.33 out of 9). Three important questions arise: (1) How likely is it that the change in help scores was due to chance? (2) Can we determine how widespread the reduction in helpfulness would be within the general population? (For instance, could we say that 10% of people who believe in determinism would be less helpful? 15%? 35%?) (3) And even if a substantial number of people would be less helpful, *how* less helpful would they be? Slightly less helpful? Moderately? Extremely?

With respect to the first question, because of a p-value of 0.03, the researchers concluded it was highly unlikely the lower average from the determinism condition was due to chance alone. To help us shed light on the second and third questions, we can use the data to determine two sorts of effect sizes: an unstandardized (or absolute) effect size and a standardized effect size. The unstandardized effect size in this case is the difference in means between the two groups. The difference in average help scores between the two groups was .9 (on a 9 point scale). Given the low p-value, this result suggests a belief in determinism (or disbelief in free will) is effective in causing an arguably substantial reduction in willingness to help (reflected in the on

¹⁰⁴ The similar help score averages from the free will and neutral conditions led the researchers to conclude that a belief in free will was already common among participants in the neutral condition. This is consistent with the finding from a 1998 survey conducted by the International Social Survey Programme that more than 70% of people—across thirty-six countries—agreed with the statement that their fate is in their own hands.

average 10% lower help score). But this tells us virtually nothing about the variability within the groups. For instance, an on average 10% reduction in help score could be the result of everyone in the determinism condition each reporting a help score of 10% less than scores reported in the control group. However, the same average difference would result from *half* of the participants in the determinism condition reporting a help score of 20% less than the scores reported in the control group and the *other half* of those in the determinism group reporting *the same score* as their counterparts in the control group.

To get a better sense of the variability of scores in the experimental group, we can appeal to a standardized effect size, e.g., Cohen's d , in which the average mean difference is couched in terms of standard deviation.¹⁰⁵ This standardized effect size will give us a better idea of the degree of variance in the help scores of the experimental condition. In general, the larger the d , the larger the absolute effect size and the smaller the variance in the population. Cohen's d in this case was 0.64, which on the standard interpretation means the effect size is moderate-to-high.

Son & Clore (2002) suggest that a willingness to help in hypothetical scenarios predicts a willingness to help in actuality; however, other researchers including West & Brown (1975) consider the relationship to be tenuous. In light of this, Baumeister et al. conducted a second experiment that measured subjects' commitment to providing *actual* help. They employed the "well validated" Katie

¹⁰⁵ Cohen (1988) defined ' d ' as the difference between two means divided by the standard deviation. According to Cohen, the standard interpretation of d values is as follows: a d value around 0.8 means the effect size is large, a d value near 0.5 means a moderate effect size, and a d value around 0.2 means a small effect size.

Banks helping paradigm (Batson et al., 1997; Maner et al., 2002). Subjects listened to a phony radio broadcast (which they believed to be authentic) during which they heard the fabricated story of Katie Banks, a college student whose parents recently were killed in a car accident. Katie asks listeners of the broadcast for help, without which she would have to drop out of school to take care of her siblings for whom she is now solely responsible. Participants were asked to indicate for how many hours (0 to 9 or more) they were willing to help Katie. Before this, subjects completed the Free Will and Determinism (FAD) scale created by Paulhus & Margesson (1994) to assess the strength of their respective beliefs in determinism and free will. The results of experiment 2 converged with the results of the first experiment. The results revealed that “disbelief in free will predicted a lower number of hours for which participants volunteered.”¹⁰⁶ In their statistical analysis, they found a significant β value of -0.30. A β value (or Beta coefficient) of -0.30 means for every standard deviation increase in the independent variable (disbelief in free will), there is a 0.3 standard deviation decrease in the dependent variable (hours of help offered). The p-value in this experiment was < 0.03 , meaning it was very unlikely the (on average) fewer hours of help offered in the determinism condition was due to chance alone.

In a third experiment, psychologists measured the effect of a belief in determinism and a disbelief in free will on aggression. As was the case in the first experiment, participants were put into free will or determinism conditions (although there was no neutral condition this time). In assessing aggression, researchers

¹⁰⁶ Baumeister et al. (2009), p. 265.

adapted work from Lieberman et al. (1999) by giving the subjects an opportunity to serve spicy food (hot salsa on tortilla chips) to someone averse to such foods (or so the subjects were led to believe). In addition, they were told these people would be required to eat all of the food served. Aggression was assessed by measuring milligrams of salsa served. Researchers found those primed to believe in determinism served more salsa than those in the free will condition. The average amount of salsa served in the determinism condition was 17.8 milligrams (standard deviation = 16.3). The average amount served in the free will condition was 9.4 milligrams ($SD = 11.6$). These results suggest a strong belief in determinism causes one to act more aggressively than someone with a belief in free will.

Because the p-value was 0.01, the experimenters concluded it was extremely unlikely the difference in outcomes was due to chance. While the results don't definitively tell us precisely how widespread the reduction in helpfulness would be among the general population, we can at least calculate Cohen's d , which in this experiment was 0.59, which on the standard interpretation means the effect size is moderate.

2.4 Stillman and Baumeister (2010): Belief in Determinism and Reduced Learning from Guilt

Stillman and Baumeister (2010) hypothesized a belief in determinism disrupts important learning that occurs from experience. They theorized people who believe in determinism are less likely to consider alternative ways of acting. They used the

same procedure employed by Baumeister et al. (2009), that is, they put random subjects into determinism, free will, and neutral conditions and had them read booklets of fifteen statements relevant to their condition for fifteen minutes. Subjects were then asked to write about a time they hurt someone. This prompt had been used by Baumeister et al., (1994, 1995) to provoke feelings of guilt. To better ensure participants had a specific event in mind, each were asked to write down the initials of the person they hurt, as well as the approximate time of day the event occurred. Subsequently, participants were asked to report the degree of guilt they felt from the event from 1 (no guilt at all) to 7 (extreme guilt). Then after being prompted to describe the event in detail, they were asked, “Do you feel you learned anything from this event?” on a scale from 1 (learned nothing at all) to 7 (learned a great deal).

The researchers found those in the determinism condition reported learning less from guilt inducing experiences as compared to those in the neutral and free will conditions. In the neutral condition, the average learning score was 6.6 ($SD = .89$), while in the determinism condition, the average score was 3.83 ($SD = 2.56$). Because the p-value was 0.048, the experimenters concluded it was unlikely the difference in the experimental and control group was due to chance. Cohen’s d in this case was 1.45, which on the standard interpretation means the effect size is very high.

In a second experiment, the researchers performed a variation of the first. They used the same booklet readings for the determinism and neutral conditions (there was no free will condition this time). They also had subjects reflect and write details about a time they hurt someone and report their feelings of guilt on the same 1

to 7 scale. One of the important differences of the second experiment was that instead of having participants self-report the degree to which they learned from guilt, the psychologists had them articulate what they learned in a written essay. The researchers believed that being able to write a meaning essay about what one learned from the guilt of hurting someone was “a more robust measure of learning than self-ratings.”¹⁰⁷ Four research assistants then rated the essays on three dimensions. Dimension one: value of the lesson learned, from 1 (not at all valuable) to 7 (extremely valuable). Second dimension: the extent to which participants seemed intent on changing their behavior, from 1 (no intention to change behavior) to 7 (very much intends to change behavior). Third dimension: the researchers rated the degree to which the lesson learned would be beneficial over an extended period of time, from 1 (not at all beneficial in the long run) to 7 (very beneficial in the long run).

The results of the second experiment converged with those from the first: deterministic beliefs reduce learning from guilt. The results of the second experiment indicated the diminished learning goes beyond the self-reports of the participants. While the researchers did not find a significant effect from induced beliefs in determinism on intentions for behavioral change, their ratings indicated the lessons from those in the determinism group were less valuable and less likely to contribute to one’s long-term well being than the subjects in the neutral condition. The average total score (i.e., the sum of the four scores assigned from the research assistants) for some participants in the determinism condition with respect to the value of the lesson

¹⁰⁷ p. 954.

learned was 16.40 ($SD = 4.83$), compared to an average total score of 23.40 ($SD = 2.07$) in the neutral condition. The p-value was 0.018, meaning it is unlikely this difference was due to chance, and Cohen's d was 1.88, meaning the effect size was high.

With respect to the long-term benefit of the lesson learned, the average total score for some participants in the determinism condition was 15.80 ($SD = 4.38$), compared to an average total score of 22.20 ($SD = .35$) in the neutral condition. The p-value was 0.032, meaning it is unlikely this difference was due to chance and Cohen's d was 2.06, meaning the effect size was very high.

2.5 Stillman et al. (2010): Belief in Determinism and Reduced Satisfaction and Productivity in the Workplace

According to two studies conducted by Stillman et al. (2010), a weak belief in free will was discovered to be a strong predictor for poor job performance and decreased job satisfaction. In the first study, subjects whose answers on a questionnaire indicated a weak belief in free will reported significantly less job satisfaction than subjects whose answers indicated a strong belief in free will. In the second study, supervisors of subjects with weak free will beliefs reported significantly poorer job performance than the reports of supervisors of subjects in the control group. If a belief in determinism causes a why-bother mentality and exports one's locus of control, it is not surprising this would negatively impact job satisfaction, which would in turn undermine performance.

But should we be as worried about the results from the Stillman et al. (2010) study as we should be about the experiments discussed earlier? Have the researchers established a *causal* relationship between a weak belief in free will and poorer job performance and decreased job satisfaction? Or have they merely established correlation? In these two studies, there was no controlled laboratory setting and the experimenters did not induce in the subjects the causal factor in question, in this case a weak belief in free will. Hence, the results are inherently weaker (than an experiment would have been) due to the increased chance of extraneous variables. That being said, the researchers did strengthen their results by assessing “potentially related variables,” including locus of control, Protestant work ethic, conscientiousness, personal vitality, and life satisfaction. This was accomplished by employing a *hierarchical multiple regression*—an HMR. This method is often used when multiple independent variables are assessed. In the first study, the dependent variable was expected workplace satisfaction. The independent variables were locus of control, Protestant work ethic, conscientiousness, personal vitality, life satisfaction, and belief in free will. Using a hierarchical multiple regression is helpful in isolating the predictive significance of a particular independent variable *over and above* other potentially related variables. In short, HMR is used to get the other related variables out of the way.

The HMR in the first study was conducted in two steps. During step one, locus of control, Protestant work ethic, and the predictive power of other potentially related independent variables were measured. In the second step, the predictive

power of the disbelief in free will was measured to see if it *added* to the predictive power of the dependent variable above and beyond the independent variables measured in step one. The experimenters found disbelief in free will *did* predict workplace success beyond the contributions of the other variables. Roughly, this means if I had an external locus of control, a Protestant work ethic, and was conscientious, I would likely succeed at work. If you subsequently strengthened my belief in free will, my workplace success would increase by a statistically significant amount. Put another way, because the researchers saw a jump in correlation data by factoring in the belief in free will influence *subsequent* to factoring in the influences of the other independent variables, they had a good reason to believe that factor alone carried predictive power. In the researcher's words, factoring in a belief in free will "significantly improved the model" such that belief in free will "was both a strong and independent predictor of expected job performance."¹⁰⁸ The p-value in this experiment was 0.006, which led the researchers to believe that it was very unlikely the difference in outcomes between the groups was due to chance alone. To be clear, however, while this study establishes correlation and predictive power, it doesn't demonstrate a causal relationship between a belief in determinism and workplace satisfaction and productivity. This is not to say philosophers should disregard the findings—just that they should be taken with a grain of salt. More attention should be given, then, to the results from the previously discussed experiments, as they provide better evidence for causal relationships.

¹⁰⁸ Stillman et al. (2010), p. 45.

If I've done what I've set out to do, it should be clear what determinism is, that there is a distinct possibility belief in determinism (and a disbelief in free will) could soon be widespread, and as a result there would be negative social ramifications. Let's talk about what can be done.

2.6 Mitigation of the Consequences: Convince Public Determinism is False or that Compatibilism is True

Perhaps the most obvious way to prevent negative consequences would be to convince the public determinism is false or that some form of compatibilism is true. I will say a little bit about how this might be accomplished. One of the ways I have in mind is quite extreme (i.e., widespread censorship). I want to be clear, however, that in the end I will not endorse such severe tactics (unless things got *really* dire). Instead I will propose what I consider to be a much more practical means of mitigating the negative consequences. For now let us consider other measures—at least one of which is rather extreme.

If negative consequences stem from a belief in determinism, couldn't we convince people that determinism is false—or at least that it is not definitively true? Adequate determinism has firm support in the literature but is by no means bulletproof. Experts in physics, psychology, and philosophy could be called to testify at a congressional hearing. If there were disagreement among them (and this isn't hard to imagine), then we would have something to tell the public that might at least shake a strong belief in determinism. Adequate determinism is a view held by many.

However, there is hardly consensus on the matter. Recall from chapter one, some believe adequate determinism is false with respect to human actions because they claim the brain is a chaotic system. Bishop (2002) addresses this possibility. He says that if the human brain is a chaotic system, quantum events could be amplified, such that adequate determinism would be false with respect to the mental states responsible for action. Whether or not this is so, Bishop notes, is highly controversial. And while the burden of proof is on those supporting this view, it is by no means a settled matter. If those with a strong belief in determinism are exposed to theories like these, which counter the thesis that all human actions are adequately determined, it's reasonable this would diminish their confidence in determinism.

Perhaps measures could be taken to prevent acquisition of belief in the first place. That is, even if determinism were true, there may be ways to prevent the public from learning about or believing it. A book burning (or file deleting) campaign could be initiated to rid the world of all information about determinism. Institutions of learning could forbid it from being taught in schools. Philosophers and members of the scientific community could be censored. A task force could be formed to discredit anyone who publicly endorsed determinism. Technology or medication could be developed to eradicate the belief from our minds. A utilitarian argument could be made: the utility gained from prevention of social unrest greatly outweighs the utility lost from censorship. We could employ the statements used to induce a disbelief in determinism from the Vohs & Schooler (2008) experiment as

propaganda. They could be put in television ads, billboards, internet banners, and fortune cookies. Anything to inculcate the public.

As long as there is freedom of speech, freedom of the press, and widespread Internet access, completely stomping out belief in determinism is likely an exercise in futility. But it may be enough to confine the believers in determinism to the minority. If the social implications of a belief in determinism got so bad, governments of the world could urge top physicists and philosophers to publicly denounce any endorsement of determinism, or at least claim that the theory is on shaky footing. If it becomes a large enough threat to homeland security, governments may have legal grounds to take measures to mitigate and prevent the belief.

Or perhaps we would not have to be so dishonest. Governments could inform the public that although determinism is likely true, everyone is encouraged to take whatever means necessary to dispel the belief. This could prove difficult at first, but over time may be possible. The argument is simple enough for anyone to understand: if too many people believe in determinism, the fabric of society could unravel. Present them with enough evidence to believe this and they may accept with open arms the government and privately funded public service announcements, billboards, speeches, etc. An ad campaign of such magnitude may be costly, but as long as that cost is offset by the utility of social preservation, the general public may be on board. And one benefit of propaganda is even when on an intellectual (or in this case a philosophical) level one knows the message to be false, it still substantially impacts one's psychology. I wouldn't be surprised if most people don't consider themselves

to be particularly persuaded or impacted by advertisements, in general. (I would probably count myself as such a person.) But advertisers don't spend cash at a rate of eight million dollars a minute on Super Bowl ads for nothing.

A law that forbade people from promoting, discussing, or believing(!) in determinism would probably not work. Much more effective would be an accord or global summit, where nations came together and made agreements in everyone's best interest. Our inherent drive toward instant gratification combined with lack of foresight has contributed to the endangerment of our ecosystem, but we try to follow the resolutions from global summits aimed at curbing our polluting habits because it's in the world's interest. Similarly, it's human nature to pursue the truth and unveil cover-ups, but in the interest of social functionality, we might follow the demands of global summits aimed at dispelling belief in determinism.

What we are talking about here some might count as a *noble lie*. In *The Republic*, Plato argued there were cases in which a lie is justified, cases in which there was some noble need.¹⁰⁹ Plato argued it was okay to mislead people into believing they couldn't transition into a higher social class by hard work or any means within one's control because such a belief, if commonplace, would disrupt social order. While Plato's example is dated (and a bad one), the general idea behind the noble lie is intuitive enough: It's okay to lie, mislead, etc. if there would be dire consequences otherwise, (in both the Plato and determinism cases the dire consequence is social disorder).

¹⁰⁹ Book 3, 414e–15c.

Here's a noble lie many would find acceptable: lying to a dying person about their chances of survival. In *The Impossible*, a film depicting the survival story of a family caught in the aftermath of the 2004 Indian Ocean Earthquake, one of the protagonists is close to death, lying on a hospital bed in an overcrowded clinic just after the disaster. Her leg is badly injured, but she tells her eleven year-old son that as long as her leg is still red, things are okay (she is a physician). At several points in the film, she asks the son about the color of her leg (she is too weak to look for herself). The first few times the son is happy to report the red coloring of her leg. The last time she asks, he looks at it with concern and says, "It's still red." Clearly the leg was turning black, but the son thinks it best to lie to his mother. The reasoning is something like: if she is going to die anyway, what's the point of burdening her with the news? Or, if she maintains an optimistic outlook, this may positively impact her physiology such that it would increase her chances of survival. This is akin to the benefit of the lie in the determinism case: lying brings a benefit.

I'm sure there are some philosophers who shudder at the thought of the 'noble' lie. We should also consider the potential worry from epistemologists who would claim that endorsing a false belief is philosophical heresy. However, if the threat to social order is substantial, the noble lie may be something we should keep on the table. Furthermore, the extent of the cover-up, propaganda, public service announcements, etc. need only be proportionate to the threat of social disharmony. So if that threat is minimal, then the censorship campaign needn't be far-reaching, which should be comforting—at least a bit—to our epistemologist friends.

If the campaign to dispel belief in determinism fails, there are still other options. Instead of trying to convince the public determinism is on shaky ground or that it is false, we could try convincing them determinism is consistent with free will (even if it turns out that it isn't). The best way to do this I can think of is to make public the case for compatibilism in a way that is clear and accessible. At least we could try convincing people there aren't conclusive counterarguments to the most promising varieties of compatibilism. We could inform the public about cutting edge compatibilist theories, including ones from Frankfurt, Strawson, Dennett, and others. In the first chapter, I presented these accounts and addressed major objections they face (see section two). And now, as promised, my alternative.

2.7 Mitigation of the Consequences: Inducing or Strengthening Alief in Free Will

If we can't prevent belief in determinism by convincing people that determinism is false or that some version of compatibilism is true, then I propose we should employ methods to induce in them an *alief* in free will (I explain what aliefs are at the beginning of the following paragraph). If inducing aliefs mitigates the negative consequences, it would be very valuable for researchers to investigate ways to induce and strengthen them (even in cases in which an opposing belief is held). I will discuss three ways that *might* work to accomplish this. One of my purposes here, then, is to lay the groundwork for how to induce or strengthen free will aliefs—further research would have to be conducted. I will argue that inducing a strong alief

in free will—or strengthening an already existent alief—will mitigate the negative behavior resulting from a belief in determinism or disbelief in free will. I will now explain what an alief is, discuss what it would mean to have an alief in *free will*, explain how in general aliefs might be induced (Gendler claims they can be), explain how an alief in free will might be induced, and, finally, discuss how a strong alief in free will would mitigate the negative behavior resulting from a belief in determinism or disbelief in free will.

‘Alief’ is a term coined by Gendler (2008a) used to denote a mental state she argues is an important part of human psychology that does a better job of explaining perplexing belief-discordant behavior than do traditional or status quo models of the mind. “Any theory that helps itself to notions like belief, desire and pretense needs to include a notion like alief in order to make proper sense of a wide range of otherwise perplexing phenomena,” Gendler writes.¹¹⁰

Roughly, an alief is a mental state with representational, affective, and behavioral (R-A-B) content automatically triggered by stimuli in one’s environment. Gendler presents four cases to motivate the need for a concept like *alief*. I will run through them too, as doing so is perhaps the best way to explain what aliefs are.

2.8 Gendler’s 4 Cases of Alief

Case 1

¹¹⁰ Gender (2008a), p. 641.

In one of the paradigm examples of an alief, Gendler references a *New York Times* article about a thrilling recent addition to the Grand Canyon: The Grand Canyon Skywalk, a glass walkway 4,000 feet above the bottom of the canyon. Visitors are required to wear special booties so not to scratch the glass below, protecting the illusion of walking on air. Rothstein, the author of the *Times* piece writes:

One woman, her left hand desperately grasping the 60-inch-high glass sides and the other clutching the arm of a patient security guard, didn't dare move toward the transparent center of the walkway. The words imprinted on the \$20 souvenir photographs taken of many venturesome souls herald completion of a daredevil stunt: 'I did it!!!'.¹¹¹

What is going on in the mind of this daredevil? Gendler argues that she clearly *believes* the walkway to be safe, evidenced by the fact that she voluntarily chose to walk it. Anyone who possessed even “a scintilla of doubt” wouldn't get near it, Gendler claims.¹¹² What is puzzling, then, is the behavior and affects of the woman; they don't seem to be inline with her belief of safety. She *believes* she's safe, yet, if you didn't know any better, you'd think she *believes* she is in danger because of her behavior. To explain the belief-behavior mismatch, Gendler argues we need the notion of alief. The woman believes she isn't in danger, but she *alieves* something quite different, something like, “Really high up, long long way down. Not a safe place to be! Get off!!”¹¹³

¹¹¹ Edward Rothstein, “Skywalk Review: Great Space, Glass Floor-Through, Canyon Views.” *The New York Times*, May 19, 2007.

¹¹² Gendler (2008a), p. 635.

¹¹³ *Ibid.*

Aliefs have representational, affective, and behavioral content and cause a set of response patterns automatically triggered by stimuli in one's environment. The stimulus in this case is the sight of the bottom of the canyon thousands of feet below. Her visual system produces the following representational content: really high up, long long way down. This generates fear and anxiety in the woman, and muscle contractions that yield—among other things—shaky footsteps.

Case 2

There are many other examples of belief-behavior mismatch. A second class Gendler borrows from Rozin, Millman, & Nemeroff (1986), who produced some interesting findings. The researchers found that many participants were hesitant to wear a freshly laundered shirt of someone they disliked. Others resisted eating soup from a brand-new bedpan. Some refused to eat high quality fudge just because it was in the shape of dog feces. Still others were less accurate when throwing darts at photos of faces of friends than at pictures of people they disliked.

What is going on in the minds of these participants? Surely, Gendler thinks, they believe the shirt is clean and that the soup and fudge aren't tainted. Yet you might think otherwise from their behavior. Why do they hesitate when they know the items are harmless? Because, Gendler claims, they possess an alief with content like: "Filthy object! [representational] Contaminated! [affective] Stay away! [behavioral]." ¹¹⁴

¹¹⁴ Gender (2008a), p. 636.

Case 3

In the third sort of case, Gendler, recounts a personal experience. She left her wallet at home, including her ID, the morning of a domestic flight. At check-in, expecting to be turned away, an airline representative simply wrote “NO ID” on her boarding pass and told Gendler she needed only to go through extra security to board. After she arrived in Baltimore (for an APA program committee meeting), Gendler asks a friend to borrow cash for the weekend. The friend obliges, and after she accepted a substantial amount of cash she searches through her bag and says, “let me just stash this in my wallet.” Surely she believed her wallet was hundreds of miles away. Yet Gendler claims she believed “Bunch of money. Needs to go into a safe place. Activate wallet-retrieval motor routine now.”¹¹⁵

Case 4

In “Fearing Fiction,” Kendall Walton (1978) described the following scenario:

Charles is watching a horror movie about a terrible green slime. He cringes in his seat as the slime oozes slowly but relentlessly over the earth destroying everything in its path. Soon a greasy head emerges from the undulating mass, and two beady eyes roll around, finally fixing on the camera. The slime, picking up speed, oozes on a new course straight towards the viewers. Charles emits a shriek and clutches desperately at his chair.¹¹⁶

Both Walton and Gendler claim that Charles believes that the slime monster isn’t real and that he is in no immediate danger. However, while Charles believes he is safe,

¹¹⁵ Gender (2008a), p. 637.

¹¹⁶ p. 5.

sitting in a movie theater, he believes (says Gendler), “Dangerous two-eyed creature heading towards me! H-e-l-p...! Activate fight or flight adrenaline now!”¹¹⁷

In all these cases, the beliefs of the individuals are clear: the skywalk is safe, the food is harmless, the wallet is nowhere near, the slime isn’t real, Gendler maintains. The individuals would even bet on their beliefs. For instance, ask the woman on the Skywalk, “If you had to bet, would you bet the skywalk is safe?” Her answer would undoubtedly be “Yes,” Gendler presumes.

Perhaps these are examples, not of alief, but of self-deception, uncertainty, or temporarily forgetting, Gendler advances. Each show up in cases of belief-behavior mismatch. Someone may believe, for instance, her spouse is cheating on them but might represent the situation to others and herself as if she believed the opposite, resulting in a ‘belief-behavior mismatch’. But Gendler doesn’t think the skywalker or Rozin’s subjects, for example, are acting frightened or out of disgust because they are deceiving themselves. Also, both Rozin’s participants and the skywalker wouldn’t hesitate to explicitly endorse their beliefs, which is what the worried spouse is much less likely to do: she doesn’t want to explicitly admit her spouse is unfaithful. It’s a painful subject, which leads to her self-deception.

Perhaps the hesitation of the skywalker and Rozin participants is similar to hesitation that can result from uncertainty, e.g., hesitation to carry an umbrella because of uncertainty about whether or not it will rain. Cases like these are strikingly similar, Gendler thinks, to the alief cases: both involve hesitation. In the

¹¹⁷ Gender (2008a), p. 637.

Skywalk case, for instance, a woman hesitates because of the view of the chasm thousands of feet below. The hesitation manifests as behavior, too, looking at an umbrella and scratching one's head in one case and shaky footsteps on the glass walkway in the other. This behavior doesn't match belief. The shaky footsteps don't align with the belief that the walkway is safe, nor does the scratching of one's head align with the belief that it won't rain.¹¹⁸ But there is an important difference, Gendler thinks. While there is uncertainty in the rain case, there is none in the Skywalk case because there is no doubt about the safety of the walkway. If the person was uncertain, then when she exited the Skywalk, she might feel relieved and say something like, "Phew... I'm glad that skywalk didn't collapse." But people who exit the Skywalk don't say things like this (at least not seriously.) The person who is uncertain about the rain might be relieved when it doesn't and could easily say something like, "Phew, I'm glad it didn't rain."

If not because of self-deception or uncertainty, perhaps the belief-behavior mismatch is due to temporarily forgetting. When I'm about to bite into the dog feces-shaped fudge, perhaps I hesitate because—for just a moment—I forget it's harmless fudge. Perhaps I grab the arms of my seat in the theater because I have temporarily forgotten the green slime isn't real. But Gendler doesn't think this could be "the full story." At least it isn't the case in the Rozin experiments, in which subjects still hesitate to, e.g., eat the soup from the sterile bedpan all the while "vividly and occurrently" entertaining the idea that the bedpan is sterile. In the wallet case,

¹¹⁸ In Gendler's example, the person avers that "it's not going to rain" but is uncertain she's right.

Gendler reports vividly and occurrently entertaining the thought that the wallet was hundreds of miles away. The skywalkers don't forget for a moment that the walkway is safe. If they did their behavior would be much more dramatic than acts of hesitation, Gendler argues.

2.9 Gendler's Characterization of Alief

Now that we have run through the four cases and competing hypotheses, we are in a good place to discuss Gendler's characterization of alief. When the woman walks on the glass skywalk thousands of feet above the Grand Canyon, her visual system suggests she's about to walk off a cliff. This stimulus activates affective response patterns and motor routines. In this case the affective response pattern includes a feeling of anxiety. The motor routines include muscle movement linked to hesitation and retreat. Her alief, then, is a mental state with representational, affective, and behavioral content that in the case of the Skywalk includes "the visual appearance of a cliff, the feeling of fear, and the motor routine of retreat."¹¹⁹

Here are Gendler's own words on *alief*:

A paradigmatic alief is a mental state with associatively-linked content that is representational, affective and behavioral, and that is activated—consciously or nonconsciously—by features of the subject's internal or ambient environment. Aliefs may be either occurrent or dispositional.¹²⁰

¹¹⁹ Gendler (2008a), p. 641.

¹²⁰ *Ibid.*, p. 642.

In paradigmatic cases, an activated alief has three sorts of components: (a) the representation of some object or concept or situation or circumstance, perhaps propositionally, perhaps non-propositionally, perhaps conceptually, perhaps nonconceptually; (b) the experience of some affective or emotional state; (c) the readying of some motor routine.¹²¹

Let's apply these words to the aliefs in the other cases. Dog feces-shaped fudge (Rozin case): visual input: sight of shape of fudge. Representational content: *This looks like dog-feces*. Affective content: feelings of disgust. Behavioral content: hesitation to pick up object and put into mouth. Wallet case: visual input: seeing a big wad of cash. Representational content: *I'm holding a big wad of cash*. Affective: feeling of responsibility, feeling protective. Behavioral: muscle movement associated with wallet retrieval. Slime case: visual input: seeing images on a movie screen. Representational: *There's a slime monster coming to get me*. Affective: fear. Behavioral: muscle movement associated with grabbing chair, fight or flight behavior.¹²²

Gendler claims we need the notion of alief to satisfactorily explain the belief-behavior mismatch in the four cases and others like them. An important part of this explanation is the fact that, in general, a person can consistently believe P and *alieve* not-P. She can believe the Skywalker is safe, but *alieve* it is not; believe the fudge is harmless, but *alieve* it is not; believe the wallet is near, but *alieve* it is not; and believe the slime monster isn't real and *alieve* that it is.

¹²¹ Gendler (2008a), p. 641.

¹²² Now that we've discussed the concept of an alief, I want to note the reason for the term 'alief'. As Gendler (2008a), p. 641 says, it's because "alief is *associative*, *action-generating*, *affect-laden*, *arational*, *automatic*, *agnostic* with respect to its content, shared with *animals*, and developmentally and conceptually *antecedent* to other cognitive attitudes."

It shouldn't be surprising, then, why appealing to Gendler's *alief* could be helpful, given our determinism/free will belief conundrum. I argue one could believe there is not free will and *alieve* that there is and that a strong alief in free will would mitigate the consequences of a belief in determinism or a disbelief in free will.

If the experiments show there is a distinct possibility that a belief in determinism (or disbelief in free will) causes enough bad behavior that we have something to worry about, what can we do about it? I claim that inducing a strong *alief* in free will (or strengthening an already existent alief) will mitigate the negative effects of a *belief* that free will doesn't exist or a belief in determinism.

Could one believe in determinism, believe there's no free will, yet *alieve* in free will? In the Skywalker case, the woman believes the walkway is safe but *alieves* it is not, which means her visual system triggers the following content: representational: *I'm really high off the ground and not supported*, affective: feelings of fear and anxiety, and behavioral: muscle contractions linked to shaky footsteps, etc. How would an alief in free will be effective in the experiments? Consider the Katie Banks experiment from Baumeister et al. (2009). How would an alief in free will cause someone with a disbelief in free will to help Katie more than they otherwise would? I imagine the situation could unfold as follows. The participant's belief in free will is weakened (perhaps even destroyed) by reading the statements from the determinism condition. Although she doesn't believe in free will, imagine the participant *alieves* in free will, that is, she has a mental state with representational, affective, and behavioral content activated by stimuli in one's environment. The stimulus in this

case is listening to the Katie Banks radio broadcast. We can imagine this triggering the following content: representational: *I'm hearing the voice of someone who needs my help*. Affective: feelings like pity, sympathy, empathy, guilt, etc. Behavioral: muscle contractions linked to moving one's hands to mark a score from 1 to 7.

What could the alief look like in the aggression experiment from Baumeister et al. (2009)? Similarly, the participant's belief in free will is weakened (or possibly destroyed) by reading the statements from the determinism condition. While she doesn't believe in free will, let's imagine the participant *alieves* in free will. The stimuli in this case is the voice of the experimenter as she explains about the salsa and the person who has to eat it. Imagine this triggers the following content: representational: *I'm hearing a voice explaining that I have the opportunity to serve spicy salsa to someone who doesn't like it*. Affective: feelings like empathy. Behavioral: muscle contractions linked with hesitation of serving salsa.

Finally, what would the alief look like in the second learning experiment from Stillman and Baumeister (2010)? Again, the participant's belief in free will is weakened (or even destroyed) by reading the statements from the determinism condition. Although she doesn't believe in free will, let's imagine the participant *alieves* in free will. The stimuli in this case is the voice of the experimenter as she prompts the participant to write about a time she hurt someone. Suppose this triggers the following content: representational: *I'm hearing a voice explaining I'm to write about a time I hurt someone*. Affective: feelings of guilt. Behavioral: muscle contractions linked with writing.

Participants in the neutral condition—and people in general—alieve in free will, I think. Part of the nature of aliefs is automaticity. Pretty much anyone can relate to the automatic response of the skywalker, the dog feces-shaped fudge eater, the wallet loser, and slime avoider. Equally relatable, I contend, are free will aliefs. Consider the plethora of everyday situations where they crop up. When one replies to a text-message, looks at a lunch menu, opens a blank word-processing document, arrives at a fork in the road. These stimuli automatically trigger representational, affective, and behavioral content. I submit that participants in the determinism condition had free will aliefs before reading deterministic statements that were then destroyed or weakened upon reading the statements. What we should do to mitigate the negative consequences of a belief in determinism or a disbelief in free will is to either induce a belief-*discordant* alief in free will or strengthen an already existent one. If we could induce or strengthen an alief in free will, i.e., trigger an R-A-B response pattern similar in kind to the pattern triggered when we receive text-messages, look at lunch menus, open blank word-processing documents, arrive at forks in the road, etc., then we should expect some mitigation of the negative behavior.

2.10 Can Aliefs be Induced?

Is the induction or strengthening of alief something that has already been done? Gendler thinks so. She argues that aliefs can be experimentally induced, as is

the case in the ‘concept priming’ experiments of Bargh et al. (1996) involving scramble sentence tests. Participants were placed into three conditions: neutral, politeness, and rudeness. Those in the neutral conditions unscrambled words to form sentences containing only neutral terms. In the politeness condition, subjects were asked to unscramble words to form sentences with terms associated with politeness. And finally, those in the rudeness condition unscrambled words to make sentences with words with rude connotations. Subsequently, participants were instructed to go into the hallway and find the experimenter for their next task, only to find the experimenter talking to another ‘subject’ (who was actually a confederate). The researchers found that participants in the rudeness group interrupted the hallway conversation earlier than those in the neutral group and that the subjects in the politeness group interrupted the conversation later than the neutral group, if at all.

Gendler argues that an alief was induced in some or all of the participants in the politeness and rudeness conditions. Specifically, for the participants in the rudeness condition, she says:

...what Bargh and his colleagues have done, I want to argue, is to induce in their different sets of subjects different sorts of occurrent alief. As the result of the pre- or quasi-conscious activation of the cluster of affective tendencies and behavioral repertoires associated with the notion of rudeness, subjects in the third condition find themselves more likely to act in ways that they would act in the presence of rudeness.¹²³

The Bargh experiment is a case of conceptual priming. This priming induced a rudeness alief. In other words, the ‘pre- or quasi-conscious’ priming of words with

¹²³ Gendler (2008a), p. 655.

rude connotations activated a group of affects and behaviors linked to the concept of rudeness.

The same explanation goes for other widely-publicized Bargh experiments, Gendler argues, such as a similar sentence scrambling test, in which some subjects unscrambled words to form sentences with terms associated with the elderly and others formed sentences with neutral terms. The participants who made sentences with elderly-associated words took more time to walk down the hall toward the elevator after the completion of the experiment. Gendler argues this is because an alief was induced with roughly the following content: “Old. Tired. Be careful walking to that elevator.”¹²⁴ Gendler argues that aliefs are induced in other experiments, as well, such as those conducted by Aarts & Dijksterhuis (2003), in which priming subjects with the image of a library card led to quieter speech and in which priming subjects with the image of an elegant dining room or the scent of soap led to fastidious eating.

2.11 Ways to Induce Free Will Aliefs

If aliefs can be induced in these cases, I see no reason why they can’t be induced in the free will case. If a simple word scramble test is enough to prime an alief, thereby affecting emotions and behavior, then we shouldn’t be surprised if there are easy ways to induce alief in free will. To induce free will alief, I suggest the

¹²⁴ Gendler (2008a), p. 657.

following: reading movie lines or watching films that are freedom inspiring, meditation on seemingly free experiences, and being mindful of deterministic and free will self-talk. I will now go into detail about each of these methods.

2.12 The Appeal to Texts, Film, and Memory

One possible way to induce free will alief would be to subject those suffering from the negative consequences to media such as film, television, and literature that are ‘freedom inspiring’, that is, media depicting a struggle between determinism and free will but that in the end suggest humans have free will. The following lines from recent films serve as good examples.

“You make yourself what you are. You have control of your own destiny.” -*The Tree of Life*, 2011.

“There are two ways through life—the way of nature and the way of grace. You have to choose the one that you will follow.” -*The Tree of Life*, 2011.

“You are fully capable of deciding your own destiny. The question you face is, ‘Which path will you choose?’. This is something only you can decide.” -*Star Trek*, 2009.

“I’ve lived for a very long time and the one thing I learned—fate doesn’t decide everything. People get to choose.” -*Hancock*, 2008.

“What you get is what you get. What you do with what you get, that’s more the point.” -*City of Ember*, 2008.

“Our choices are what make us who we are. We always have the choice to do what’s right.” *Spiderman 3*, 2007.

“The future has not been written. There is no fate but what we make for ourselves.” -*Terminator 3: Rise of the Machines*, 2003.

“It is not our abilities that show what we truly are, it is our choices.” - *Harry Potter and the Chamber of Secrets*, 2002.

Other films include: *Gattaca*, *Minority Report*, and *Groundhog Day*.

The lines could be printed on a card carried in one's wallet or placed somewhere frequently glanced (a computer desktop image, above the front door, hanging from a car rearview mirror). A smartphone app could be engineered to display the lines in a compelling way. Instead of just the lines, the films themselves (or clips thereof) could be employed. Clips from different films could be used individually or edited together into compilations. Professional filmmakers could be hired to ensure the videos are as gripping and compelling as possible.

Why would reading certain texts or watching movies strengthen or induce an alief in free will? Think of them as a form of conditioning or priming. Looking down at the bottom of the Grand Canyon triggered in the woman on the Skywalk an alief with the following content: the visual appearance of a cliff, the feeling of fear, and the motor routine of retreat. The fear is a healthy one that crops up for evolutionary reasons, Gendler notes. But imagine the woman had no such alief and had feelings of overconfidence and behaved erratically. In similar potentially dangerous situations (like sitting at the top of a Ferris wheel), suppose she has comparable emotions and also behaved erratically. Couldn't we condition the woman to associate a healthy amount of fear and trepidation with these sorts of potentially dangerous environments? Wouldn't one way to do this be to repeatedly show her films (fictional or documentary-style) depicting careless people being seriously injured or killed after falling from great heights? The films in the free will case

would serve the same purpose, that is, serve to condition the desired affects and behaviors and mitigate those predicted by the experiments.

At what frequency should the texts and videos be viewed? They could be read or viewed for short intervals (e.g., 5-15 minutes) at varying frequencies (e.g., weekly, daily, multiple times per day, etc.). Should people be required to watch them or should they only be viewed on a voluntary basis? That depends on the severity of the problem. If the negative behavior severely threatened social functionality, e.g., if the economy started to significantly weaken because of job dissatisfaction and low productivity because of a widespread disbelief in free will, then it might be prudent for governments to mandate screenings of the media. Or perhaps, they would only need to be employed on a case-by-case basis for those suspected of suffering from the negative behavior. They could be part of correctional, rehabilitation, or probation programs or could be used as a prescription or treatment by a doctor or psychologist. Support groups like alcoholics or gamblers anonymous could use them. They could be implemented in the public education system or as part of training in the workplace. There many ways the media could be employed; my aim here is simply to unveil some of the possibilities.

If the films weren't effective, then let us encourage those individuals to focus on past actions they *considered* free. For instance, what they chose to eat for lunch yesterday or last summer's spontaneous decision to take a trip to Rome. Or we could encourage them to meditate on all the seemingly possible actions they could engage in in the moment (like moving a finger or limb, standing up, getting something to eat,

doing something creative or random, writing a song or poem, getting in a car and driving to some place they've never been before). Or we could encourage them to meditate on the difference between doing any of aforementioned and times when they didn't or wouldn't feel free (e.g., being imprisoned, being held at gunpoint, being at the in-laws, being blackmailed, coerced, etc.).

2.13 Self-Talk

Another therapy would be to monitor intrapersonal communication, specifically 'self-talk'. A person with a strong belief in determinism says things to herself like, "I have no free will," "My behavior is controlled more by external than internal factors," "I'm a cog in a machine," "Why bother with effort and acting 'morally'", etc. I will now draw an analogy between these assertions and the self-talk typical among individuals who believe themselves to be physically old.

In the old age case, the problematic strong belief is a belief in physical old age. Common to this belief is negative self-talk, e.g., "I'm past my prime," "My life is over," "I'm ugly," "I'm not as valuable anymore," "I'm irrelevant and obsolete," "I can't do anything without help," "I'm worthless," etc. There are many negative emotions and behaviors that may follow. Someone who believes herself to be biologically old may feel: depressed, irrelevant, despair, helpless (against the passage of time), and have morbid feelings and low self-esteem. Her behaviors may include: an inactive lifestyle, weight gain or loss, undergoing cosmetic plastic surgery,

purchasing a sports car or motorcycle, taking hair-loss drugs, taking erectile dysfunction medication, suicide, etc.

In the old age case, individuals who engaged in positive self-talk or avoided negative self-talk are more likely to avoid the negative behavior. I believe the negative behavior in the free will case can be similarly mitigated. I maintain engaging in positive self-talk and avoiding negative self-talk in the free will case would induce a free will alief or strengthen an already existent free will alief by conditioning or priming the triggering of an R-A-B response pattern similar to the patterns triggered in individuals not engaging in pernicious behavior. (That is, the response patterns we see in individuals who pick an entrée from a menu, decide how to reply to a text message, pick a path from a fork in the road, etc.)

Volumes of studies have been produced on the effects of positive and negative self-talk. The results aren't surprising. Negative self-talk is correlated with an increase in negative behaviors and emotions and a decrease in positive behavior and emotions. Positive self-talk is correlated with reduced negative emotions and behavior and increased positive emotions and behavior. Moreover, positive self-talk is effective in triggering desired actions, increasing effort, controlling attention, reducing anxiety, providing self-reward, etc.¹²⁵

Theodorakis & Chroni (2001) investigated the effect of self-talk on the accuracy of shooting baskets. Participants were randomly assigned to different groups: some to the control group and some to the 'relax' group. In the first trial,

¹²⁵ According to Hardy, Jones, & Gould, (1996).

subjects in both groups were instructed to shoot for three minutes from five different positions (all 4.5 meters away from the basket). In the second trial, participants in the control group were given the same instructions; however, subjects in the relax group were told (in private) to say the word 'relax' before each shot. The average number of shots taken by a member of the control group was 53.05 ($SD = 5.09$) and the average number of successful shots was 20.22 ($SD = 5.83$). The average number of shots taken by a member of the relax group was 53.40 ($SD = 4.76$) and the average number of successful shots was 23.90 ($SD = 4.81$). The p-value was less than 0.05, so it's unlikely the difference in results was due to chance alone. Cohen's d-value was 0.69, indicating a fairly high effect size.

In a similar experiment, Van Raalte et al. (1995)'s results indicated that positive self-talk improved the accuracy of dart throwing and that negative self-talk diminished it. In the experiment participants were randomly divided into three conditions: positive self-talk, negative self-talk, and control. Each member of all the groups was told to make fifteen attempts to hit the bull's-eye of a dartboard as many times as they could. After each throw, the distance from the bull's-eye was measured in centimeters. Any miss of the board entirely was counted as 17.75 cm away from the bull's-eye (the radius of the board). Participants in the positive self-talk group were told to say, "You can do it," before each attempt, while those in the negative self-talk group were prompted to say, "You cannot do it," before each throw. Control group subjects weren't prompted to say either. The distance from the bull's-eye for each throw was added together for each participant (e.g., 3.4 cm + 6.5 cm + 17.75 cm,

etc.). The average ‘total miss score’ for the three groups were as follows: positive self-talk group: 69.75 ($SD = 22.58$); control group: 84.75 ($SD = 24.48$); negative self-talk group: 92.75 ($SD = 25.88$). Cohen’s d value for the effect size of the positive self-talk on accuracy compared to the control group was calculated at 0.64, which is moderate-to-high.

Being mindful of self-talk may help one throw darts or shoot baskets, but that doesn’t mean it will mitigate all the negative consequences of a disbelief in free will, one might claim. There may be some cases where being mindful of self-talk just wouldn’t work—the Vohs & Schooler (2008) experiment, for instance, in which participants cheated by paying themselves for wrong answers on a test. These participants had a compelling reason to forgo being mindful of self-talk because forgoing it entailed a monetary reward. A sixty-five year old might forgo being mindful of self-talk when it comes to, say, social security benefits or senior discounts. The worry is that self-talk therapy may be effective to a certain extent but not when there are strong forces—like the lure of cash—tempting them to forgo it.

However, I don’t think these cases are analogous. In the old age case, let us suppose that being mindful of self-talk serves to instill in a person the alief that she is young. It’s not clear to me the alief would destroy her belief that a sufficient number of years have passed to enjoy social security benefits, nor would it undermine her motivation to pick up a social security check. Suppose in the cheating case the positive self-talk serves to instill in a person an alief in free will. Possessing an alief

in free will *would* undermine one's motivation to pay themselves for wrong answers—or at least it would *likely* undermine this motivation, I contend.

Alieving that you're young means that while you may be physically old, you have a positive attitude about your age and often behave like a young person. Joan Rivers, Betty White, Bill Murray, and George Burns are examples of celebrities who I consider to possess (or who did possess) an alief in youth. Burns once said, "You can't help getting older, but you don't have to get old."¹²⁶ Similarly, possessing a free will alief means that while you may be physically determined, you have a positive attitude about your actions and you often behave as if you are morally responsible for them. Most people who have never heard of determinism—sans a few who suffer from addiction, depression, etc.—possess free will aliefs, I maintain. If George Burns alieved he was young until the day he passed, I can envision how this alief in youth would combat feelings of despair, irrelevancy, helplessness and behaviors like an inactive lifestyle, taking hair loss drugs, etc., but it is not obvious that it would prevent him from collecting social security (let's assume he needed the money). I don't see as clearly, however, a case of someone who alieves herself to be free (say, as the result of being mindful of self-talk) who steals by paying herself for wrong answers on a test. The alief—as I see it—would sap her motivation to cheat. The following is an example of how this might happen.

Imagine George Burns is a participant in the determinism group that has the opportunity to steal by paying themselves for wrong answers on a test. Suppose, like

¹²⁶ Epstein (2011), p. 174.

many of the others in his group, he *does* pay himself for wrong answers after reading the deterministic statements. After the experiment, George practices being mindful of self-talk. He catches himself thinking and saying things like, “I’m a cog in a machine that is the universe” and “I don’t have free will” and replaces those statements with ones like, “I’m in control of my future” and “What I do in life is up to me.” After weeks of this training, George forms the alief that he is free. Not long after this, George participates in a similar experiment (though he is unaware of the similarity; George thinks it’s an experiment about memory). Because of his alief in free will, Burns has a positive attitude about his determined actions and often acts as if he is morally responsible for them. Once again, George reads the deterministic statements presented to the participants in the experimental group. He does believe in determinism, but he also alieves in free will (as a result of the self-talk training). Is it easy to envision George stealing by paying himself for wrong answers on a test a second time? I think not—at least it’s not obviously easy to envision. And let’s remember that even if I’m wrong in this case, that is, even if a strong alief would *not* be enough to prevent cheating, this does little to undermine my claim that the inducing of strong free will aliefs would mitigate the other negative consequences like increased aggression and reduced helping, job satisfaction, productivity, etc.

Appealing to the suggested therapy—the reading of statements and quotations, the watching of freedom inspiring films, the meditating on seemingly free actions and possibilities, or the monitoring of self-talk—would induce or strengthen an *alief* in free will, I maintain. As automatic responses to stimuli, it stands to reason that the

strength of aliefs is subject to classical conditioning. The suggested therapies could be part of such a conditioning regimen. Alief is connected to the experience of an affective or ‘emotional state’, says Gendler. The appeal to inspirational quotations and compelling cinema is designed to habituate into the subject the emotions associated with freedom, especially during times she questions whether or not she is. In general, the therapies should serve to inculcate in the affected an alief that generates representational, affective, and behavioral content, such that her behavior is just as moral as a person with a genuine belief in free will.

I will conclude by saying a little bit about whether or not I think a particular negative consequence would be better mitigated by the proposed therapies than the others. (And let’s focus on the negative consequences for which I provided detail, i.e., reduced helpfulness, increased aggression, reduced learning from emotions, and reduced satisfaction and productivity in the workplace.) For instance, would watching freedom-inspiring films be more effective in mitigating reduced helping more than it would in mitigating increased aggression? I won’t go as far to say that it’s clear to me that a particular negative effect would be mitigated more or less than any other, given a specific therapy. This is an empirical claim that is best left to the researchers. That being said, if forced to choose one of the therapies, it might be most prudent to begin with being mindful of self-talk, as there is data suggesting this is effective in triggering desired actions, increasing effort, controlling attention, reducing anxiety, providing self-reward, as Hardy, Jones, & Gould, (1996) suggest. Although, I wouldn’t be surprised if the viewing of freedom-inspiring films would in

general better mitigate the negative consequences than would the other therapies, if only because of the power and salience of cinema.

I have argued that belief in determinism will be problematic in the not too distant future. The results of the experiments and studies I scrutinized suggest a strong belief in determinism leads to reduced helping, aggression, reduced learning from emotions like guilt, and less job satisfaction and productivity. If we are unable to convince people that determinism is false or that they should believe in free will anyway, then I propose we induce in them free will aliefs. I have attempted to support the claim that strengthening or inducing alief in free will would mitigate the negative behavior. Even if I have not fully succeeded, I hope I have at least demonstrated the need for more empirical research on whether free will alief can be induced and whether or not this induction would help combat the free will disbelief problem. If this is the case, I have provided—at the bare minimum—a direction for valuable research.

3. Free Will₂ and the Law

There is a comic from *The New Yorker Magazine* depicting a defendant pleading his alleged innocence to a judge. The caption reads “Not guilty by reason of genetic determinism, Your Honor.” The joke is about genetic determinism, but the caption could have just as easily read, “Not guilty by reason of *causal* determinism.”¹²⁷ If determinism is true, then all defendants—not just ones in comics—may have a substantial defense. If being guilty of a criminal offense requires that one could have done otherwise, then—assuming determinism to be true—no one would be guilty of *any* offense because no one could have done otherwise than she did.¹²⁸

3.1 Assumptions of the Criminal Law

Does the law require one to have the ability to do otherwise to be guilty of a criminal offense? According to Michelle Cotton, yes: An assumption of the criminal law is that “persons can be held responsible for their actions because they have freely chosen them, rather than had them determined by forces beyond their control.”¹²⁹ And Cotton writes that freely choosing to commit a criminal offense means “the

¹²⁷ Or more specifically, “Not guilty by reason of adequate scientific determinism.” I discussed the details of adequate scientific determinism in chapter one and at the beginning of chapter two. As in the previous chapter, I will usually use ‘determinism’ as a shorthand for ‘adequate scientific determinism’.

¹²⁸ At least not under U.S. law, but presumably the same implication would follow for most or all legal systems.

¹²⁹ Cotton (2005), p.1.

defendant could have done other than he did.”¹³⁰ If she is right, the criminal law works under the assumption that individuals can only be held responsible for acts for which they could have done otherwise. But no one could have done otherwise, if adequate determinism is true. Thus—following the letter of the law—no one can be held responsible for a criminal offense. If no one can be held responsible for an offense, then no one can be found guilty of an offense. One could get away with murder... literally.¹³¹

3.2 Free Will₂

To avoid legal calamity—or at least to eliminate this loophole—I suggest the criminal law incorporate a concept I call *free will*₂.¹³² I propose the law be amended such that a person is guilty of a criminal offense if a jury of her peers unanimously judged that she freely₂ committed the offense, whereas:

X is an act of free will₂ if and only if *X* would be judged to be an act of free will₁ by a person who is clear-thinking, unbiased, rational,

¹³⁰ Cotton (2005), p.1.

¹³¹ Or at least the killing of someone. ‘Murder’ might not be the right term, if there’s no free will₁.

¹³² Free will₂ is distinct from traditional free will, what I call ‘free will₁’. (I will also use terms like ‘free₁’, ‘freely₁’, and ‘freedom₁’.) Someone performs an act of free will₁ if and only if (a) the person wills it, i.e., the act fulfills one of the person’s desires, (b) the act isn’t constrained, and (c) the person performing the act could have done otherwise. Here is some background on these conditions. (a) is hinted at in the literature at least as early as Hume. According to Hume, if a person has free will, she must possess “a power of acting or of not acting, according to the determination of the will” (Hume 1748/1977). There are various ways in which acts can be constrained. A.J. Ayer (1954) does a good job of cataloging them into *internal* and *external* constraints. Examples of internal constraints include gambling addiction and OCD. External constraints include, e.g., being manipulated at gunpoint or being physically forced to do something, e.g., being removed from a building by security. The (c) condition shows up in the literature at least as early as Hume (2000) and is covered in detail in Nowell Smith (1948). These three conditions comprise what I call ‘the traditional conception of free will’.

dispassionate, consistent, and well-informed (sans awareness of any reason to question the existence of free will₁).¹³³

And the process of jury selection should mirror the process in the current U.S. Justice System.

Under this revision, what exactly would juries be asked to do? And who, precisely, is this clear-thinking, unbiased, rational, etc. person? Under current U.S. law—at least according to Cotton—to be found guilty of a criminal offense requires the person in question could have done otherwise. If no one could have done otherwise (because of determinism), then no one can be legally found guilty of a criminal offense. We need a system under which persons can be found guilty of criminal offenses. To this end, I propose we amend the law such that legal conviction of an individual is possible, even if that person couldn't have done otherwise.

If it doesn't make sense for convictions to depend upon whether a person freely₁ committed some act (because there are no such acts on the assumptions of this chapter), then on what should they depend? Let them depend on a *judgment* that the person freely₁ committed some act, I submit. While determinism may threaten our

¹³³ There are other attributes we could require of the person, of course. I don't intend for this list to be unamendable. In an earlier draft, I included the attributes of being compassionate, empathetic, and kind. On the recommendation of one of my committee members, I omit them here to avoid the tension between being *dispassionate* and *compassionate*. The person ought to be dispassionate, rational, consistent, and unbiased if she is to avoid judgments that are unfair, inconsistent, biased, and that are made in haste or in the heat of the moment. However, since arguably there are times a good judge or jury member would display mercy toward a defendant, a case can be made this person should also be compassionate, empathetic, and kind. But if we can't have it both ways, that is, if there is no way to reconcile the tension between these seemingly inconsistent traits, then I submit it would be better to keep the attributes I include in my analysis of free will₂ (above) than the 'affective' attributes I omitted. There is a lot of literature on the tension between dispassion and compassion and whether or not compassion and similar affective attributes have a place in the courtroom. For more on this, see *Emotion and the Law: Psychological Perspectives*, edited by Bornstein and Wiener, especially, chapter 3: "Emotional Influences on Judgments of Legal Blame: How They Happen, Whether They Should, and What to Do About It," by Neal Feigenson.

being free₁, it doesn't necessarily threaten *judgments* that a person is free₁. But, as I suggested in the second chapter, the number of people who judge acts as free₁ may be shrinking because increasingly people are believing in determinism.

For this reason, *free will*₂ depends on whether or not an act would be judged by a person who (besides having sound judgment) is *unaware* of any reason not to believe in free will₁. Jury members would be asked to discuss whether or not—in their best opinions—a person who (a) possesses sound judgment (i.e., is clear-thinking, unbiased, rational, dispassionate, and consistent), (b) is well-informed about the facts of the case,¹³⁴ and who (c) is unaware of any reason to disbelieve in free will—would judge that the defendant acted freely₁. For short, let's call this hypothetical person 'Jim'. Jim is someone who (a) possesses sound judgment, (b) is well-informed, and who (c) *mustn't* be aware of any reason to not believe in free will₁.

If it were known the jurors possessed a strong belief in free will₁, it would be unnecessary to instruct them to imagine what Jim would say; we could just ask them if *they* judge that the defendant acted freely₁. If this were the case, then we would expect the judgments of both the free will₁-believing jurors and Jim to converge, especially since the courts require jurors to use good judgment and to be impartial. However, in the near future some jury members may not believe in free will₁, which may lead them to judge a defendant's act as *unfree*₁. Convictions based on the judgments of these jury members may prove difficult. As a preventative measure, we could instruct all jury members (whether or not they believe in free will₁) to vote not

¹³⁴ Except for the fact that the person's act was determined.

on whether *they* would judge the act in question as free₁ but whether *Jim* would.

One might worry that it would be asking too much of our jurors to figure out what Jim's judgments would be. How are they supposed to know about the judgments of a hypothetical person with such-and-such qualities? Firstly, if the juror believes in free will and is using her best judgment (as is required by the court), then—as I just mentioned—her judgment is likely to converge with Jim's would-be judgment. Secondly, if the juror believes in determinism (and as a result does not believe in free will₁), then to figure out how Jim would judge the act of the defendant, we could instruct the jury member to reflect upon how she would judge the defendant's act *before* developing a disbelief in free will₁. (Perhaps the disbelief formed in college upon taking classes in philosophy, physics, and neuroscience.) Finally, if the juror never possessed a belief in free will₁, then she should ask herself: "How would an impartial person judge the defendant's act before there was any talk of determinism or the problem of free will?"

Jury members *already* are often required to disregard evidence, testimony, publicity, etc. For instance, suppose in a murder case a jury hears a witness' testimony that is very condemning of the defendant. But later it comes out that the prosecution improperly coached the witness.¹³⁵ Judges in this case would deem the testimony inadmissible and instruct the jury to disregard it. Moreover, jurors are often required to disregard their exposure to 'pre-trial publicity' that is prejudicial, as was the case in the trials of O.J. Simpson and the officers involved in the beating of

¹³⁵ An attorney telling a witness that it is okay to 'bend the truth' under oath is a common example of witness coaching.

Rodney King.¹³⁶ Since the Fifth Amendment guarantees any defendant the right to not testify, judges often instruct juries not to infer anything from a defendant declining to testify.¹³⁷

Since there presently exists a mechanism for ensuring that juries will disregard inadmissible information, implementing my proposed amendment may not be as difficult as one might think. If determinism threatens the functionality of criminal justice systems, let us employ this mechanism to ensure juries disregard determinism. In their consideration of the would-be judgments of Jim, they would be doing just this, i.e., disregarding determinism. What, precisely, would be the instructions to the jury? The judge might say to the jurors, “Do your best to pretend the defendant’s act wasn’t causally determined. That is, pretend you have never heard of determinism, and given the facts of the case use your best judgment to determine if the defendant freely performed the act in question.” Or something like, “Imagine a hypothetical juror named Jim. Jim has never heard of determinism and strongly believes in free will. Also, Jim is clear-thinking, unbiased, rational, dispassionate, consistent, and is aware of the same facts of the case that have been made available to you. In your best opinion, would Jim judge the defendant’s alleged criminal act to be free?” I contend requiring jury members to consider the would-be judgments of Jim is not requiring too much; it is not requiring something beyond their imaginary powers. Such a requirement is no more demanding than what is already asked of actual juries, including a common instruction to disregard inadmissible

¹³⁶ Lieberman and Arndt (2000), p. 678.

¹³⁷ Hemmens, Brody, & Spohn (2013), p. 256.

evidence to which the jury has already been exposed.

On the assumptions of this chapter, it's true that defendants of all stripes could plead to the judge, "Not guilty by reason of determinism," and be correct. Taking my suggestion, however, the judge could scoff at the reply and say: "You are quite correct under previous law you would not have been found guilty. But the law has been revised such that 'the determinism defense' will no longer exonerate you. Under this amendment, convictions could be made, thus avoiding social upheaval (or at least closing the legal loophole).

3.3 Does the Law Need Amending?

Two important questions a skeptic may ask. Does the law need such an amendment? That is, is it really threatened by determinism? And, secondly, would such an amendment be satisfactory? Perhaps, it's true my amendment would give the court a rule under which convictions could be made. *But would the rule be just?* If not, my amendment may close a legal loophole, but not in an acceptable way. That is, the amendment may give the courts *legal* grounds to make convictions, but not *moral* grounds.

Here is a dialog capturing this potential problem.

JUDGE:	I'm sorry Ms. Kelsch, your daughter, Meggie, has been sentenced to ten years in federal prison.
MS. KELSCH:	On what basis, Your Honor?
JUDGE:	She was convicted of murder in the first degree, which under the new law means that a jury of her peers unanimously judged that she freely ₂ murdered

someone.

MS. KELSCH: I know what it means to freely₁ murder someone; it means someone committed murder and among other things had the ability to do otherwise.

JUDGE: That's right. And because all of our actions are determined, no one is free₁ because no one has that ability.

MS. KELSCH: If no one freely₁ does anything, then isn't my daughter legally innocent?

JUDGE: Before a recent revision to the law, she would have been innocent. But under the new law you can be convicted if you are judged to have freely₂ committed a criminal offense.

MS. KELSCH: What does it mean to be judged to have freely₂ committed a criminal offense?

JUDGE: In this case it means a jury of her peers unanimously judged your daughter's act to be a act of free will₂.

MS. KELSCH: Go on...

JUDGE: That is, it was the judgment of the jury that your daughter's act would be judged as free₁ by a person with sound judgment (i.e., someone who is rational, unbiased, clear-thinking, dispassionate, and consistent.)—someone who was unaware of determinism or any reason to not believe in free will₁.

MS. KELSCH: I see. And I also see the need for such a revision. Because of adequate determinism, everyone would be innocent of any crime because no one has free will₁. But society couldn't function under such a legal system—or at least there would be a loophole in the law. What would deter people from committing crimes if they knew they couldn't be convicted? Their sense of morality?

JUDGE: An astute observation.

MS. KELSCH: So the legislators needed a way around the problem. Thus they defined 'guilty of a criminal offense' in a way that didn't require the defendant to have free will₁. Clever.

JUDGE: That's right.

MS. KELSCH: But here's my problem. I understand that if my daughter murdered someone but had the ability to do otherwise, that many—though perhaps not myself!—would say she deserves to be convicted. But I don't think a reasonable person would say Meggie *deserves* to be convicted under your new gerrymandered

definition of ‘free will’. So what if her peers judged her act would be judged to be free₁. Why does this mean she deserves to be put away for ten years?

JUDGE: You would not be the first to make such a complaint.

Mrs. Kelsch’s complaint captures what I take to be the main worry about my proposal. Even if we have closed ‘the loophole’ in the law and can *legally* justify conviction, such a revision to the law wouldn’t necessarily be justified in an intuitive sense. If there were a law that every March 22nd all first born children are to be tossed out to sea, that would give the government legal grounds to kill children, but it certainly wouldn’t provide moral grounds. We are faced with two important questions: Does the law need to be amended in light of determinism? And is the amendment I propose satisfactory? I will first address the former.

Belief in determinism is anything but new. Theories of causal determinism can be traced back at least as early as the presocratics. Scientific determinism, grounded in the early eighteenth century work of Newton, was codified by Laplace in 1820. And more sophisticated versions of scientific determinism like D1, D2, D3 (see chapter one), etc. have appeared in the literature for years. One may ask: Why bother revising the law if it has been effective for centuries in which determinism was common knowledge. That is, hasn’t the criminal law ostensibly worked just fine, despite widespread knowledge of determinism? One might question the need for such an amendment if courts seem adept at ignoring determinism.

Perhaps some day in the not too distant future, a local trial court judge took ‘the determinism defense’ seriously. Suppose it were in part because of the impeccable rhetoric of a young, hot-shot attorney. Word might spread quickly of this

feat. Imagine the defense team of an ultra affluent celebrity client (someone like O.J. Simpson, Michael Jackson, Martha Stewart, or former Illinois Governor, Rod Blagojevich) hires the young hot-shot. The case is heard by a State Supreme Court of Appeals. The determinism defense prevails. It makes national news. Eventually the validity of the determinism defense is considered by the U.S. Supreme Court. The defense's main argument rides on the expert testimony of the world's premier psychologists, physicists, and philosophers, who argue that as a matter of fact no human could have done otherwise than she in fact did. They claim this fact in conjunction with the definition of 'guilty of a criminal offense' under current U.S. Law entails neither their client *nor anyone* can be legally convicted of any criminal offense. To claim otherwise would be a blatant injustice and a gross misinterpretation of the law. Suppose on a 5-4 ruling, the court finds the celebrity defendant innocent on all counts due to the fact that he couldn't have done otherwise. The presiding justice's opinion sets a new precedent: no person can be convicted of any crime because in order for the conviction to be valid, it would require that an individual possess an ability no one has, namely, the ability to do otherwise.

Imagine the crime that would ensue, when the world learned of this unprecedented ruling. Imagine the chaos and repercussions for the rest of the world's legal systems. There would be a scrambling, no doubt, on behalf of Congress to amend the law. I suggest they use the concept, *free will*₂.

Not convinced such a story could become reality? Would the courts ever take seriously 'the universe made me do it' defense? In the Summer of 1924 in what

Chicago newspapers touted as the ‘Trial of the Century’, Clarence Darrow¹³⁸ appealed to deterministic thinking in his defense of Nathan Leopold, Jr. and Richard Loeb, who were both accused of the kidnapping and murder of fourteen-year-old Bobby Franks.¹³⁹ In a twelve-hour closing argument, Darrow persuaded Judge John Caverly not to sentence Leopold and Loeb to be hanged. Instead they both received a life sentence in prison, despite public pressure for the death penalty. At the time, Leopold and Loeb were both eighteen-year-olds who came from well-known, affluent Chicago families and were both very intelligent. Loeb was the youngest on record to graduate from the University of Michigan, and Leopold was a law student on his way to Harvard Law. In his closing argument Darrow, an open proponent of determinism, said this about Richard Loeb:

What had this boy to do with it? He was not his own father; he was not his own mother; he was not his own grandparents. All of this was handed to him.... He did not make himself and yet he is to be compelled to pay.¹⁴⁰

Darrow’s implication was the two young men were not morally responsible for the killing. The blame ought to be placed on causes preceding their existence, their parents, grandparents, etc. Elsewhere Darrow asked:

Why did they kill little Bobby Franks? Not for money, not for spite; not for hate... They killed him because they were made that way. Because somewhere in the infinite processes that go to the making up of the boy or the man something slipped...¹⁴¹

¹³⁸ Darrow is also well known for his role in the Scopes ‘Monkey’ Trial, in which he defended the right of John Scopes to teach evolution in Tennessee public schools. The play and novel *Inherit the Wind* are based on this trial.

¹³⁹ Payment (2004).

¹⁴⁰ Lief, Caldwell, Bycel (1998), p. 188.

¹⁴¹ *Ibid.*, p. 173.

These words are in the same spirit of the preceding quotation. Humans don't have free will₁ and cannot be held responsible because we are part of a causal chain producing unavoidable events, Darrow argued.

This sort of defense has been effective in recent years, too. In 2002, a defense team argued their client, Thomas Koskovich, on trial for the murder of two pizza delivery men, did not deserve the death penalty because of environmental factors beyond Koskovich's control—such as the abuse from and later abandonment by his parents and his addictive personality. (Allegedly, the day his mother left she told Koskovich she “loved him the least” of her children.) The judge was persuaded, and Koskovich received a life sentence, avoiding the death penalty.¹⁴²

According to Cotton, ‘Deterministic thinking’ has increasingly been a part of the criminal law. “By the [twentieth] century’s third quarter, the criminal law had accepted a significant amount of deterministic thinking in virtually every one of the areas in which the issue had arisen: the insanity defense, related and analogous defenses, expert witness testimony on mental state, juvenile justice, and sentence mitigation,” she notes.¹⁴³ Allowing this sort of thinking into the criminal law may be the first step in opening Pandora’s Box. The more we allow it, the closer we are to traversing down a “slippery slope, at the end of which we would have nullified the entire criminal law.”¹⁴⁴ Or as Rebecca Dresser puts it, “The feared outcome would be a system in which no one could be held responsible for any of the crimes they

¹⁴² Mansnerus and Hanley (2001).

¹⁴³ Cotton (2005), p. 3.

¹⁴⁴ Meir Dan-Cohen (1983), “Actus Reus.” In S.H. Kadish, ed., *Encyclopedia of Crime and Justice* (New York: Free Press).

commit.”¹⁴⁵ If deterministic thinking begins to saturate the criminal law, then as Cathy Gere suggests, the determinism defense would be a ‘get-out-of-jail-free card’ for all.¹⁴⁶

To avoid this chaos, I suggest the courts and legislatures prepare to amend the criminal law. I propose incorporating the concept *free will*₂ is a first step in this amendment process. I am not alone in claiming the criminal law is in need of amendment in light of determinism. Cotton argues that “courts and legislatures have tended to see any acceptance of determinism... as requiring fundamental systemic revision, if not the need for a different system entirely.”¹⁴⁷

The courts and legislatures understand that *if* determinism is true, a fundamental revision is likely to be necessary. But *do* the courts accept the thesis of determinism? Cotton claims that many see the acceptance as inevitable. “The law should also accommodate determinism because courts generally suspect that determinism will one day supersede free will as the prevailing account of criminal behavior,” she says.¹⁴⁸ She cites Judge Posner’s worry that someday we will be able to attribute all criminal behavior to factors beyond their control.¹⁴⁹ Since the acceptance of determinism by the courts and legislatures is a real possibility, I agree

¹⁴⁵ Dresser (1999), p. 170. Although ‘crimes’ may not be an appropriate term, if no one has free will₁. Perhaps we could use the term, ‘illegal offenses’, as a substitute.

¹⁴⁶ Gere (2011), p. 2.

¹⁴⁷ Cotton (2005), p. 4.

¹⁴⁸ *Ibid.*, p. 44.

¹⁴⁹ *United States v. Beserra*, 967 F.2d 254, 256 (7th Cir. 1992).

with Cotton that measures ought to be taken *now* in preparation.¹⁵⁰ She argues the courts' possible embracing of determinism is reason to accommodate determinism into the law now. I argue the concept of *free will*₂ can be a viable part of this accommodation.

3.4 Would Incorporating Free Will₂ into the Law be Morally Justified?

But do people who freely₂ commit offenses really *deserve* to be convicted? I take this to be the most salient objection to my account. My opponent may admit there is nothing stopping us from incorporating *free will*₂ into the law. But she may insist we *shouldn't* because it would mean the conviction of individuals who don't intuitively deserve to be convicted. Michael Moore argues, "A just legal system requires people to be truly responsible. If moral responsibility requires free action and if there is no such thing as free action, we cannot found our moral system on some supposed 'postulate' of free action that we believe, as a matter of scientific fact, to be false."¹⁵¹ I take him to be talking about free will₁. His argument as I see it is:

- (1) If we don't have free will₁, no one is morally responsible for their actions.
- (2) If people aren't morally responsible for their actions, then we can't have a just legal system.
- ∴ (3) If we don't have free will₁, then we can't have a just legal system.

(1) is a controversial premise, which I will end up arguing is false. Setting that aside

¹⁵⁰ Certainly, anticipating a necessary revision or amendment to current law is prudent. And even if it turns out that determinism is false, it's not clear that amending the law in the way to be proposed would be substantially harmful.

¹⁵¹ Moore (2010), p. 515.

for a moment, I want to examine a claim similar to (1),¹⁵² but where ‘morally’ is replaced with ‘legally’, call it (1*):

(1*) If we don’t have free will₁, then no one is legally responsible for their actions.

Without revising the criminal law, (1*) is true (because of the assumptions I have shown to be deeply seated in the law). But if we do revise it by, for instance, incorporating *free will*₂, then (1*) is false. (1*) would be false because we could hold people *legally* responsible, if for instance, we wrote the law such that a person is legally responsible for a criminal offense only if she freely₂ committed it. But Moore’s response would likely be that while this revision provides a rule under which we can *legally* find persons guilty and *legally* hold them responsible, it doesn’t mean these individuals truly are *morally responsible* nor that the criminal law is *morally* just. We could still convict individuals who don’t intuitively deserve to be convicted. It would not be acceptable if this were the consequence of integrating *free will*₂ into the law.

This, as I previously stated, is the biggest objection to my thesis. My thesis is that not only would the incorporation of *free will*₂ give us legal grounds to convict and punish persons, it would give us moral grounds. Accordingly, I think (1) is false. To show this, I will employ the work of five philosophers, namely: P.F. Strawson, John Fischer, Mark Ravizza, Nicolas Southwood, and Jay Wallace. Though none of these five address whether or not we should make use of a concept like *free will*₂ in the law, each offers a unique and thoughtful perspective that, as we will see, is very

¹⁵² It will soon become evident why I want to examine (1*) instead of (1).

relevant to this salient objection.

3.5 Fischer, Ravizza, and Strawson

Clearly, individuals who freely₂ commit criminal offenses would be convictable if we revise the criminal law such that ‘a person is convictable’ *just means* she freely₂ committed a criminal offense. This should be uncontroversial. The controversy is whether or not individuals who freely₂ commit criminal offenses *should* be convicted. In other words, under that revision to the law, a person would be *legally* convictable, but the question still remains: *should* we convict her? My opponent will argue conviction under this revision is unjust because persons wouldn’t be truly morally responsible for their crimes. To be truly morally responsible for a crime, a person must have *freely*₁ committed an offense, and this is impossible if adequate determinism is true. (Note, this is the contrapositive of (1)). With the help of the five aforementioned philosophers, I plan to argue we are morally justified (not just legally) in convicting persons who freely₂ commit criminal offenses. To do this, I need to show a person does not have to freely₁ commit a crime to be truly morally responsible for it. I am not the first person to postulate this. P.F. Strawson argued for this in “Freedom and Resentment.” I take Strawson’s account and the one I provide here to be closely related. They are related enough that I would hardly be surprised if Strawsonians would be sympathetic to my proposal. However, in some cases, I distance myself from Strawson and claim there are some objections (or potential

objections) to the account he provides in “Freedom and Resentment,” which do not apply to mine. The following is one of those cases.

Recall from chapter one that for Strawson the conditions of moral responsibility and free will are *constituted* by the reactive attitudes and practices that inform our judgments of moral responsibility. These attitudes are “natural human reactions to the good or ill will or indifference of others towards us,” which include resentment, anger, gratitude, and forgiveness, just to name a few.¹⁵³ One of the potential threats to this view of freedom and responsibility comes from John Fischer and Mark Ravizza, whose work I discussed in chapter one. Fischer and Ravizza claim, “Strawson’s theory may reasonably be said to give an account of what it is for agents to be held responsible, but there seems to be a difference between being *held* responsible and actually *being* responsible. Surely it is possible one can be held responsible even though one in fact is not responsible.”¹⁵⁴ Their worry is that Strawson blurs the line between being *held* responsible and *being* responsible. However, Strawson wants to break down the line between the two. He maintains individuals are responsible because they are capable of participating in interpersonal relationships in which they hold each other responsible, irrespective of the truth of determinism. Since the conditions of moral responsibility and free will are *constituted* by the reactive attitudes and practices that inform our judgments of moral responsibility, one could say being responsible for Strawson *consists in* being held

¹⁵³ Strawson (1962), p. 195. Strawson maintains that even the general acceptance of determinism would not lead to the “decay or the repudiation” of these attitudes and practices.

¹⁵⁴ Fischer and Ravizza (1993), p. 18.

responsible. Watson succinctly summarizes this position: “In Strawson’s view there is no such independent notion of responsibility that explains the propriety of the reactive attitudes. The explanatory priority is the other way around: It is not that we hold people responsible because they *are* responsible; rather the idea (*our* idea) that we are responsible is to be understood by the practice.”¹⁵⁵

Fischer and Ravizza are skeptical of this view. Further, they claim if Strawson is correct that responsibility is to be understood by our practices, and not the other way around, then his theory has no ability to *revise* or *criticize* existing practices. What if there were communities who “hold people responsible who are intuitively not.”¹⁵⁶ Can holding persons responsible truly make them so? They offer this thought experiment:

Imagine, for example, a society in which severely retarded or mentally disturbed individuals are resented, blamed, and harshly punished for their failure to adhere to the norms of the community. (Perhaps the society attributes their failure to poor character or an evil nature.)¹⁵⁷

Because every member of the community holds this attitude, Strawson may be committed to the view that the mentally disabled are truly responsible for their behavior. But this does not jibe with our intuitions, Fischer and Ravizza maintain, and they posit that any view that commits us to claiming these individuals as responsible for their behavior is mistaken.

Strawson’s account may be susceptible to such an attack (an attack I don’t necessarily endorse). Whether or not he can offer an adequate response on his behalf

¹⁵⁵ Watson (1988), p. 117.

¹⁵⁶ Fischer and Ravizza (1993), p. 18.

¹⁵⁷ *Ibid.*

is not my purpose here. Instead, I will show how my account of *free will*₂, while similar to Strawson's account of free will, is different enough to not be open to this attack. First, however, I will say more about how Strawson's and my account are similar. On both our accounts, free will depends on persons. For Strawson free will depends on the reactive *attitudes* of people. For mine it depends on the *judgments* of persons. On both accounts, there is neither responsibility nor free will₂, independent of attitudes or judgments, respectively. Neither Strawson nor I are realists in this regard. Some realists about responsibility and free will see difficulties in Strawson-like accounts. For them free will consists in something independent of judgments about free will, and responsibility consists in something independent of attitudes. One attractive feature of these realist accounts is they are not problematized by thought experiments like Fischer and Ravizza's, in which a community holds the mentally disabled responsible for their unwelcomed behavior. Realists can claim the community is not justified in holding these individuals responsible because responsibility consists in something 'deeper' than the practice of holding individuals responsible. The members of the community are wrong in virtue of this attitude-independent notion of responsibility, which "explains the propriety of the reactive attitudes."¹⁵⁸

While Fischer and Ravizza's objection may be damaging to Strawson's account, I don't think it represents a fatal blow to mine. If Strawsonians are committed to the notion that the attitudes of the community (particularly unanimous

¹⁵⁸ Watson (1988), p. 117.

ones) determine responsibility, then the objection could be detrimental.¹⁵⁹ While my account is similarly ‘mind-dependent’, its strength—at least in defending against this realist attack—is that the bedrock for *free will*₂ is not attitudes but *judgments*.

To understand the substance of this defense, consider an analogous thought experiment about free will. Suppose the community unanimously considered the mentally disabled to be *free*₁ agents. This runs counter to what is actually the case. For instance, if a mentally disabled person leaves her home unsupervised, takes the car out, and crashes into a tree, we do not blame her. Our attitudes are neither such that the person is responsible nor that she possessed *free will*₁ in the act. But suppose society were different such that we *did* hold the mentally disabled responsible for this sort of car accident and we did consider them to have *free will*₁. On my view, am I committed to admitting the mentally disabled driver has *free will*₂? No. On my account, the act is an act of *free will*₂ if and only if Jim would judge it to be an act of *free will*₁.

I don’t think the thought experiment can get off the ground because it is not possible—or at least it is exceedingly difficult—to imagine Jim with all those attributes, judging the act in question as *free*₁. If a clear-thinking, well-informed, unbiased, rational, dispassionate, and consistent person would judge the disabled

¹⁵⁹ Though I suspect Strawson could get around this objection. For instance, he could specify that the conditions of moral responsibility and free will are constituted by the *actual* reactive attitudes and practices that inform our judgments of moral responsibility, as they are now. If there were a shift in attitudes, then, conditions of morality would not shift with them. That is, Strawson could hold that only *actual* attitudes and practices of humans constitute the conditions of morality and not attitudes—*were they to shift*—as they do in the Fischer and Ravizza case where the mentally disabled are held morally responsible. This sort of reply has been offered by many defending ‘response-dependent’ accounts of moral responsibility. These accounts suggesting morality is constituted by *actual* responses of subjects in certain circumstances are sometimes called ‘rigid’ accounts. For more on this, see Wiggins (1987), Johnston (1989), and Lewis (1989).

person's crashing a car into a tree as free₁, I wouldn't find it terribly difficult to accept the judgment. If I am not alone in this intuition, then my account may survive the counterexample.

We could tell a similar story for moral responsibility. Because of the constraints on the character of Jim, I don't envision Jim judging the mentally disabled driver to be morally responsible for crashing the vehicle. On Strawson's view, the driver is responsible in light of the reactive attitudes of the community, which constitute moral responsibility. On my view, I don't think the driver would be responsible because of the difficulty in imagining Jim judging her to be free₁.

3.6 Wallace and Strawson

Strawson's theory of moral responsibility is a radical departure from the more traditional standpoint that our attitudes or beliefs about responsibility track external, mind-independent facts. However, as Jay Wallace argues, this traditional view is "wedded to the dubious image of a prior and independent realm of facts about moral responsibility," facts that are independent of "our practice of holding people responsible."¹⁶⁰ So while the traditional standpoint is, perhaps, more in line with folk intuitions, it too faces objections. Each view represents a horn of a dilemma. One horn, what I dub *the Socratic view*, is the view that moral responsibility is prior and independent of our practice of holding people responsible. The problem for this view,

¹⁶⁰ Wallace (1994), p. 89.

as Wallace suggests, is these facts are strange, dubious, and very different in nature from the sort of facts we are accustomed to—empirical facts, for instance. We can verify empirical claims, but there does not seem to be a way to verify claims about moral responsibility on the Socratic view. Not being empirically verifiable makes these ‘facts’ about moral responsibility strange. If we can explain moral responsibility without unverifiable facts, then we should, Wallace might say. Ockham certainly would agree.

On the second view or horn, which I call *the Euthyphronic view*, moral responsibility *depends* on the practice of holding people morally responsible. One salient objection, which we saw before from Fischer and Ravizza, is that if responsibility depends on the practice of a community, what are we to say when that community holds persons responsible who intuitively to us are not? On Strawson’s view there is no constraint, independent of the attitudes of a community, from which to criticize. Because the bedrock of Strawson’s account is attitudes and not judgments, his approach is considered by some, like Wallace, to be noncognitivist in character.¹⁶¹ Wallace asserts this because for Strawson “holding people morally responsible is understood not in terms of beliefs about the people who are held morally responsible, but in terms of the emotions one feels toward them.”¹⁶² Strawson’s view, then, is open to the challenges that face noncognitivists, for instance, the criticism from T.M. Scanlon that this approach to responsibility does not get at

¹⁶¹ Noncognitivism about moral responsibility is the view that sentences about moral responsibility do not express propositions (i.e., they are not truth-apt) nor do they express beliefs. Instead these sentences express non-cognitive attitudes. Thus, on this view, there are no truth conditions for moral responsibility.

¹⁶² Wallace (1994), p. 74.

what is fundamental to moral judgment.¹⁶³ Hence, Wallace claims we should accept this conclusion “only as a last resort.”¹⁶⁴ Wallace suggests the following analysis of moral responsibility, which is a departure from Strawson’s noncognitivism and is much closer to the kind of analysis I offer for *free will*₂. On this view from Wallace, facts about responsibility are “fixed by our dispositions to hold people responsible in favorable conditions.”¹⁶⁵ He calls this *D*:

D *S* is morally responsible (for action *x*) if and only if we are disposed, under favorable conditions, to hold *s* morally responsible (for action *x*).¹⁶⁶

While *D* is not exactly a Strawsonian view, Wallace considers it an improvement and a middle ground that “concerns all parties to the traditional debate,” i.e., the parties holding the two aforementioned views or horns of the dilemma.¹⁶⁷ How does *D* serve as a middle ground? For one, it is not the Strawsonian view because on *D* there are truth conditions about moral responsibility. (Recall, for Strawson, there aren’t any.) As Wallace puts it, this view is “not committed either to the noncognitivist position that claims about moral responsibility lack truth-conditions, or to the eliminativist position that there are no moral facts about moral responsibility.”¹⁶⁸ At the same time, it avoids the pitfalls of the Socratic view because the facts about moral responsibility are not independent of our practice of holding people morally responsible. These

¹⁶³ Scanlon (1988).

¹⁶⁴ Wallace (1994), p. 89.

¹⁶⁵ *Ibid.*

¹⁶⁶ The catch-all clause ‘under favorable conditions’ needs clarification. While I can’t say precisely what Wallace means, I do know that he includes it to dismiss moral judgments of persons who are—for instance—fatigued, in a highly emotional state (e.g., sorrow or elation), have a faulty memory, etc.

¹⁶⁷ Wallace (1994), p. 89.

¹⁶⁸ *Ibid.*, p. 90.

‘theoretical virtues’¹⁶⁹ are the same benefits promised by dispositional or response-dependent accounts of color and value. These accounts have been dubbed varieties of *qualified realism*.¹⁷⁰

How would a *D*-like analysis of free will (call it ‘*F*’) compare to my account of *free will*₂? Let’s compare:

<i>F</i>	<i>S</i> has free will (with respect to action <i>x</i>) if and only if we are disposed, under favorable conditions, to hold that <i>s</i> freely performed <i>x</i> .
<i>FW</i> ₂	<i>X</i> is an act of free will ₂ if and only if <i>X</i> would be judged to be an act of free will ₁ by a person who is clear-thinking, unbiased, rational, dispassionate, consistent, and well-informed (sans awareness of any reason to question the existence of free will ₁).

One strength of my analysis over *F* is specificity. My analysis offers details about what the conditions are, i.e., the required qualities of Jim. Also, under *F*, it’s not clear who ‘we’ is. ‘We’ could be an individual, a small group, a large community, all humans, etc. As I see it, my analysis is consistent with *F* yet provides more details. Wallace says in a footnote that “different ways of filling out the favorable conditions clause would yield widely divergent applications of the dispositional strategy, some of which might be more promising than others.”¹⁷¹ I think the way I fill out *FW*₂ counts as one of those promising ways.

The benefit of this approach is it “allows us to say there are facts about moral responsibility [or freedom], without postulating the complete independence of the

¹⁶⁹ Wallace (1994), p. 90.

¹⁷⁰ For more on this, see Johnston (1989) and Pettit (1991).

¹⁷¹ Wallace (1994), p. 89.

facts from our practice of holding people responsible.”¹⁷² The approach also offers truth-conditions for statements about moral responsibility. On *F* those truth-conditions depend on dispositions to hold individuals responsible, and on my analysis the truth-conditions depend on the would-be judgments of Jim. Thus, the advantage of my position and *F* over Strawson’s—and noncognitivist positions in general—is that both offer the qualified realism that avoids the pitfalls of both the Euthyphronic horns and offers truth-conditions and facts about free will.

3.7 Southwood and the Arbitrariness Worry

In the first chapter, I introduced Mulligan, Simmons, and Smith’s account of truth-makers. I will make use of their work again to highlight an important difference between *free will*₁ and *free will*₂. The truth-maker of the proposition, *Tommy freely*₁ *took a sip of water* is *the fact* that Tommy freely₁ took a sip of water. It’s vital to note the truth-maker of this fact is *not* a judgment that Tommy is free₁. If Tommy freely₁ took a sip of water, this means, among other things, he took a sip of water, but could have done otherwise. Whether or not Tommy has this ability also does not depend on a *judgment* that he could have done otherwise. The truth-maker for the proposition *Tommy freely*₂ *took a sip of water* is the fact that Tommy freely₂ took a sip of water. This fact *does* depend on judgments. Even if no one *judged* (or would judge) that you freely₁ did something—like taking a sip of water—it’s possible that you freely₁ took

¹⁷² Wallace (1994), pp. 89-90.

the sip of water. In contrast, if no one judged or would judge that you freely₁ took a sip of water, it would be *impossible* that you freely₂ took the sip. In short, free₁ actions need not be judged to be free₁, while free₂ actions necessarily would be judged to be free₁.¹⁷³ In light of this, let us say my analysis of *free will*₂ is a *judgment*-based account and that the account of *free will*₁ is a *non-judgment*-based account. The Socratic account I discussed in the previous section is a sort of non-judgment-based account.

As I've expressed repeatedly, I take it the main objection to incorporating *free will*₂ into the law is that it would allow us to convict individuals who don't intuitively deserve to be convicted. Recall Ms. Kelsch's reservation about the amendment to the law. Even if we have closed 'the loophole' in the law and can *legally* justify conviction, such a revision to the law wouldn't necessarily be justified in an intuitive sense. I think the reason for this is that it is the intuition of most clear-thinking people that a proper conviction should be made only if the freedom were non-judgment-based. What I call 'judgment-based accounts' are often thought to be unsatisfactory because the judgments on which they are based seem arbitrary. Attitude-based accounts, like Strawson's noncognitivist account of moral responsibility, may also be thought to be unsatisfactorily rooted in attitudes that are susceptible to being arbitrary. The attitudes of a group, whether it be a small

¹⁷³ However, some may take issue with this, or at least may wonder if people who believe in free will₁ are committed to the view that free will₁ obtains whether or not anyone *would* judge it to. Perhaps, then, there are fewer stark free will realists (i.e., those committed to the belief that free will₁ obtains even if no one would judge it to) than I originally envisioned. I imagine there are some, though. But I need not say all realists about free will are of this stark variety. The point I will emphasize in this section, instead, is that although I offer a 'judgment-based' account, I can satisfactorily address concerns about arbitrariness.

community or society in general are susceptible to being arbitrary because there is no attitude or judgment-independent constraint placed upon the judgments or attitudes. This problem is exemplified in the Fischer-Ravizza thought experiment in which the community's attitudes were such that they held mentally disabled persons responsible for their actions. With no attitude independent fact about moral responsibility to fall back on, one may be forced to accept the strange attitudes as the basis of responsibility. If the community holds the mentally disabled responsible, then they are. Loosely, whomever the community holds responsible *are* responsible. Sound arbitrary? That is precisely the problem.

But while Strawson's account may face the arbitrariness charge, my account is less susceptible to it. Any arbitrariness of judgments is mitigated by Jim's qualities. In effect, Jim is what Nicholas Southwood calls a 'deliberative contractor'. In his defense of contractualism¹⁷⁴ as a foundation of morality, Southwood says this about them:

Although deliberative contractors remain, in Habermas's phrase, 'the last court of appeal',¹⁷⁵ the procedural constraints that constitute deliberative rationality are sufficiently robust to ensure that there is nothing arbitrary about the validity-conferring power that they wield, even in the absence of explanatorily more fundamental considerations.¹⁷⁶

On my account Jim's judgment represents the last court of appeal. Given Jim's qualities, I think there is nothing arbitrary about the freedom₂-conferring power they wield, even in the absence of facts about free will₁.

¹⁷⁴ Moral contractualists like Southwood maintain morality depends on an agreement or contract. This thinking is rooted in the social contract theory of Hobbes.

¹⁷⁵ Habermas (1990), p. 67.

¹⁷⁶ Southwood (2010), p. 186.

3.8 Could Justice be Based on a Falsehood?

Have I shown that convictions based on the would-be judgments of Jim could be just? One critique that, perhaps, I have not satisfactorily addressed is that if we amend the criminal justice system as I suggest, convictions would be unjust because they would be based on a falsehood, namely that a defendant's act was not determined. Convictions would be based on the would-be judgments of Jim, someone who believes determinism is false (recall this is not the case on the assumptions in this chapter). One of the reasons I argued the amended system would be just is because it would allow us to continue convicting people who we would want to convict, e.g., those who have killed, stolen, inflicted violence, etc. For instance, in the current system suppose someone were convicted of the assault and battery, rape, and first degree murder of an innocent eleven year old girl. (Suppose, too, that the evidence against the defendant was devastating: there was a confession, eyewitness testimony, forensic evidence, guilty plea, etc.) Even if there were overwhelming public support for this conviction, it would be wrongful on the assumptions in this chapter. If society is to function well, then arguably we need to be able to make these sorts of convictions. If not amending the law would make it difficult (or impossible!) to make convictions, then there would be pragmatic value in an amended system that would allow for the sort of convictions we want. I claimed convictions in the amended system would be justified because of the pragmatic value

the system would carry with it. Someone who believes convictions would be unjust in a determined world—some staunch incompatibilists, for instance—may never find my proposal acceptable. But I hope to have persuaded some that the amended legal system I propose is justified in virtue of the pragmatic value afforded by it.

Perhaps there is another way to defend my proposal. Under current law, the courts often ‘ignore’ certain facts. As previously mentioned, jurors are often instructed to ignore things like inadmissible evidence, pretrial publicity, testimony that has been thrown out, etc.—despite often times being factual. Courts instruct juries to dismiss certain facts because taking them into consideration would render an unfair verdict. Why not count determinism as inadmissible evidence? We instruct jurors to ignore determinism (via Jim’s ignorant would-be judgment), just as we instruct them to ignore factual pieces of evidence, testimony, etc. that would undermine the functionality of the justice system. My critic might say we can’t ignore determinism just to get the result we want. This would be like ignoring someone’s innocence to prove her guilt. We can’t ignore evidence that a person is innocent just because it suits us. I tend to agree, but this is a special case because we can’t afford to *not* ignore determinism. Doing so would be too costly to social functionality.

It might be said that we already do this, that is, the courts already ignore determinism. Perhaps they have been, but thinkers like Cotton argue this cannot or will not continue for much longer. Recall she says “courts generally suspect that determinism will one day supersede free will as the prevailing account of criminal

behavior.”¹⁷⁷ Even if the courts have been ignoring determinism for a long time, then at least the courts should be more explicit about this. At the very least, then, my proposed amendment would serve this function.

3.9 Similar Proposals?

Have there been other proposals to amend the law in light of determinism? And how do they stack up to the project I offer here? In my research I have found no detailed proposals on how to revise the criminal law in light of determinism. However, I will discuss the work of two thinkers, both of whom offer general suggestions about how the criminal law ought to be changed because of determinism. The first suggestion comes from Christopher Slobogin. Slobogin, in “The Civilization of the Criminal,” argues a *preventive* model of criminal law ought replace the current *punishment* model of criminal justice.¹⁷⁸ The second suggestion is made by Robert Wright. Wright argues in *The Moral Animal* that we should embrace a utilitarian system of punishment in light of determinism.¹⁷⁹ Here are some details on each.

Slobogin claims given a deterministic universe we would be prudent to move toward a preventive model of criminal justice as opposed to a punishment model.¹⁸⁰

On a punishment model, the government punishes individuals for *past* actions,

¹⁷⁷ Cotton (2005), p. 44.

¹⁷⁸ Slobogin (2005).

¹⁷⁹ Wright (1994).

¹⁸⁰ A preventive model of criminal justice has also been suggested by Glueck (1962), Menninger (1966), and Wootton (1963).

according to the principle of just desert. On this principle, a person ought to be punished only if she is morally blameworthy for an action. The preventive model, in contrast, discounts the principle of just desert, and focuses entirely on preventing—or at least reducing—the ‘dangerousness’ of the convicted. A person is dangerous if she is potentially harmful to others. One of the most important reasons why we would shift toward the preventive model is because of growing support for determinism, Slobogin argues. If determinism is true, he says, then individuals don’t have control¹⁸¹ over their actions and we would be hard pressed to properly assign them moral blame because a punishment-based justice system would be invalid. However, the preventive model is well suited for a deterministic universe, Slobogin thinks, because whether or not anyone deserves moral blame, it is still prudent to prevent crimes, if only for the benefit to society. On the preventive model, the dangerousness of the individual is assessed by the state and measures are taken to reduce or eliminate the danger.

Robert Wright, in *The Moral Animal*, argues that given determinism,¹⁸² we are left with two options: we can either (a) redefine free will to ‘artificially restore’ it or can (b) give up on the notion altogether and embrace a legal system whose function is entirely utilitarian. A utilitarian justice system would be aimed at maximizing utility or happiness. Instead of retribution, measures would be taken that would benefit society overall, including rehabilitation and incarcerating a criminal as a means to

¹⁸¹ Though he doesn’t make this explicit, I take Slobogin to believe that having “control” implies having the ability to do otherwise.

¹⁸² Wright writes about genetic determinism or what I call in chapter one ‘bio-environmental’ determinism; however, the same implication follows from the truth of adequate scientific determinism.

prevent future offenses or as a deterrent. If determinism means free will is an illusion, Wright argues these measures are justified as they preserve social order. In other words, in a deterministic world, it might be that no one deserves to be punished; however, incarceration and rehabilitation are justified because the “welfare of society at large is more important than the welfare of the individual offender.”¹⁸³

I will now discuss how my proposal is similar to Wright’s and Slobogin’s, beginning with Wright’s. Although a utilitarian may assent to incorporating *free will*₂ into the law, I don’t take my proposal to align well with Wright’s (b) option because I don’t think we have to go as far as to abandon the notion of free will altogether. I do, however, see my project in a similar vein with option (a) because I think my thesis can be accurately described both as an artificial restoration and a redefinition. However, where Wright doesn’t offer more than an example of *how* to redefine the concept of *free will*, I contribute a detailed analysis. Wright gives one example of how to redefine free will. (Recall this redefinition is in light of genetic or bio-environmental determinism, which isn’t necessarily equivalent to scientific determinism.) On the definition of free will under which the courts currently work, information about the genetic makeup of a drug addict can influence the decisions of the courts. For instance, sentences in possession cases have been mitigated on account of a defendant’s “addictive personality.” If genetic (or scientific) determinism is the case, the worry is that one’s genetic makeup (or prior conditions and laws) would entail the innocence of everyone. Wright doesn’t offer a detailed or

¹⁸³ Cashmore (2010), p. 4503.

redefinition of the concept of free will;¹⁸⁴ he tells us only that such a definition would have to disentangle free will from genetics. One way would be to declare “that the existence of a biochemical correlate has no bearing on whether a behavior is volitional.”¹⁸⁵ My redefinition of the concept of *free will*, i.e., *free will*₂ does provide a detailed analysis. Not only does it disentangle free will from biochemical correlates (as well as from the ability to do otherwise), it gives precise conditions, under which *free will*₂ is constituted by the would-be judgments of a well-informed person with sound judgment.

Slobogin’s project is similar to mine in that he offers the courts a new course to take in case determinism turns out to be true. One problem, however, is that it doesn’t seem to allow for moral praise or blame. Slobogin argues moving toward a preventive model of criminal justice would ‘civilize’ the law. While we may agree non-preventive punishment has a tendency to be viewed as uncivilized (just think back to crucifixions and guillotines), leaving it entirely by the wayside may not be the way to go. Slobogin’s reasoning is valid: If determinism is true, no one is morally responsible for their actions. If no one is morally responsible for their actions, then the courts ought to shift from the punishment to the preventive model of law. Thus, if determinism is true, the courts ought to shift from the punishment to the preventive

¹⁸⁴ But perhaps Wright needn’t offer one. On Wright’s account we could maintain that determinism is true, let go of concepts like *moral responsibility*, *free will*, *justice*, etc., yet still incarcerate and rehabilitate individuals in order to maximize utility. So why should the courts bother redefining or artificially restoring free will if their only role is to maximize utility? One reason would be to maintain a legal system that more closely resembles our own. I contend the legal system I propose much more closely resembles the current one than would a purely utilitarian or preventive model. It isn’t easy to amend a legal system. For logistical and practical reasons, it makes more sense to transition to the system I have in mind than to these more radical alternatives, I maintain.

¹⁸⁵ Wright (1994), p. 355.

model of law. However, I take the first claim of this syllogism to be false. If we give an account of moral responsibility that mirrors my account of free will₂, then we can have moral responsibility in a determined world. A person would be morally responsible, then, if and only if she is judged to be so by Jim—a well-informed person with sound judgment. Jim would likely judge persons as morally responsible because Jim is unaware of any reason to question the existence of free will₁.

3.10 Incorporation of Free Will₂ into the Law

It would be prudent to offer an example of how the criminal law could take the suggestion I put forth in this chapter. The following case should suffice. Suppose, Emily, a woman with an opiate addiction, is caught by local law enforcement in Santa Cruz County with sixty grams of heroin and five-thousand dollars of stolen cash (the money she allegedly acquired by pickpocketing). Heroin possession of any amount is illegal in every state in the U.S. She is charged with possession and theft. The case goes to a local trial court, during which her attorney claims Emily is innocent of the crimes on account of her opiate addiction. Emily's physician serves as an expert witness and testifies that Emily indeed has an opiate addiction. Closing arguments are made by both Emily's attorney and the prosecution, and her guilt or innocence is decided by a jury of her peers. The story thus far is no different than how things work in the current legal system. But here the paths diverge. Instead of the jurors deciding whether or not they unanimously judge Emily's act as free₁, they deliberate

about whether or not Emily's act was free₂. That is, they discuss among each other if a well-informed person (sans awareness of any reason to question free will₁) with sound judgment (who we've named 'Jim') would judge Emily's act as free₁. The instructions to the jury might be as follows:

Though some of you might not believe in free will (i.e., you believe no action can result from the unconstrained will of a person, such that she could have done otherwise), deliberate about whether a well-informed person (sans awareness of any reason to question the existence of free will) with sound judgment (i.e. someone who is clear-thinking, rational, unbiased, dispassionate, and consistent) *would* judge Emily's act as free. Assume this person is privy to the same facts of the case that have been made available to you.

Both the jurors deciding Emily's guilt or innocence and the judge determining her sentence may be aware of the truth of adequate determinism. They may also be aware the law has been modified on account of this truth. It may be common knowledge in the courtroom the old definition of 'free' was problematic because under it convictions were not possible. Social order was in jeopardy. The law was amended to ensure convictions could be legally made. In the Emily case, this means, roughly, that her innocence or guilt and sentence would be determined by a jury of her peers and a judge who ask themselves: "in my best opinion, would a person with sound judgment who didn't know about determinism judge that Emily freely₁ possessed narcotics and stole money?"

If we incorporate *free will*₂ into the law, a person may be *legally* convictable, but *should* we convict her? Only if she is truly morally responsible for the crime, Moore and others say. Are persons who freely₂ commit crimes morally responsible for them? If Strawson is right, moral responsibility consists in attitudes not

threatened by adequate determinism. Thus, I think a Strawsonian would claim we have moral grounds to convict those who freely₂ commit crimes. My thesis has the benefit of being in a similar vein of a Strawsonian account of free will. However, it is different in that it is a kind of qualified realism, not noncognitivism; it rests on judgments, not attitudes.

There are at least two benefits from this difference. The first is my thesis allows for truth conditions with respect to free will₂. There are no truth-makers or bearers for the Strawsonian.¹⁸⁶ However on my account, there are. The truth-maker of the proposition *Tom freely₂ took a sip of water* is the fact that Tom freely₂ took a sip of water. This is a virtue in the eyes of those philosophers—from the literature there are quite a few—who believe a proper account of moral responsibility (and free will) ought to be a realists account, i.e., it must provide truth-conditions.

The second benefit of my account over Strawson's is that the judgments of those who determine free will₂ are not arbitrary.¹⁸⁷ Free will₂ is based on the judgments of a person who is clear-thinking, well-informed, unbiased, rational, dispassionate, and consistent. Nicholas Southwood might call such a person a 'deliberative contractor'. I claimed it is difficult to imagine a deliberative contractor—such as Jim—possessing 'strange' attitudes, like those of the community

¹⁸⁶ Although, one could argue I'm making Strawson's account out to be more distinct from mine than need be. Yes, there are no truth-makers or bearers on Strawson's account because *attitudes* serve as the basis of moral responsibility, which are not truth bearing (unlike judgments). That being said, the truth-makers of my account are facts like that Jim would judge that so-and-so's act was performed freely₁, which are arguably quite different than the sorts of mind-independent states of affairs that serve as truth-makers for some realist accounts of free will (such as the fact that so-and-so's act really was performed freely₁).

¹⁸⁷ Although, as I suggested earlier, it may be the worry about arbitrariness could be adequately addressed by Strawson.

who held the mentally disabled responsible for their actions and would judge their actions as free₁. If the members of a jury unanimously decide that Jim would judge an act in question as free₁, then I think a court *does* have moral grounds to convict them—even if free will₁ does not exist.

Conclusion

In this dissertation, I addressed the problem of deterministic thinking. Deterministic thinking, as we saw, is problematic for everyday social interaction, as well as for the criminal law. I argued belief in determinism is increasing in strength and number, that is, the number of people who believe in determinism is increasing, as well as the degree to which people believe. If this number and degree become too great, then as the recent findings from psychology indicate, there could be negative social consequences. I presented in brief many of these findings. In greater detail, however, I discussed the findings from Baumeister et al. (2009), Stillman and Baumeister (2010), and Stillman et al. (2010). These researchers provided evidence that belief in determinism or disbelief in free will causes increased aggression and reduced helpfulness, reduced learning from guilt, and reduced satisfaction and productivity in the workplace, respectively.

If we could convince the public that determinism is false or that compatibilism is true, that would likely mitigate the negative consequences. I discussed a few ways we might go about doing this, including making public the objections to determinism, to show the public ‘the jury is still out’. If that didn’t work, I suggested we popularize the cutting-edge varieties of compatibilism discussed in chapter one. If they could be presented in a clear and plain manner, they might catch on, thereby preventing the free will disbelief problem. But if these methods weren’t effective, that is, if the public held firm in their belief in determinism and disbelief in free will,

which led to a substantial increase in the negative behavior predicted by the studies from psychology, then I recommended we induce in them an *alief* in free will. This, I argued would be effective, because it is possible to alieve in free will while *disbelieving* in free will. Further, I argued a person with a strong alief in free will would not engage in the pernicious behavior—at least not to the same degree—because of the affective and behavioral aspects of aliefs. Aliefs can be induced, claims Gendler. So why not free will aliefs? To induce them, I proposed that people reflect on texts, films, and memories that inspire freedom. Many examples of these I provided. If this wasn't effective in inducing free will aliefs, I recommended the monitoring of self-talk. I drew an analogy between the efficacy of self-talk with respect to old age and self-talk with respect to free will.

Deterministic thinking is problematic for the criminal law because of assumptions implicit to it, at least according to some like Michelle Cotton. The law requires that anyone convictable of a criminal offense must have the ability to do otherwise (than commit the offense). Since according to many proponents of determinism, this ability can't exist in a determined world, courts may face a problem. Cotton and others argue deterministic thinking is already present in the criminal law, and some claim it is only a matter of time before a dramatic amendment or revision will be required. In anticipation of this, I provided in chapter three a way to amend the law such that convictions could be made even on the assumption that determinism is true. That is, I provided a definition of 'guilty' under which convictions could be legally made. Without any amendment or revision, I argued the

law contains a loophole, through which the innocence of any defendant would be guaranteed. Though, even if this loophole currently exists, clearly the courts have been ignoring it. However, Cotton argues that it is only a matter of time before the courts will be unable to ignore the problem. The main worry about my proposal in the third chapter is that while amending the criminal law to incorporate free will₂ may successfully close the legal loophole, it would not allow for convictions that are *morally* justified, only legally justified. I argued this is not the case because the qualities of ‘Jim’, the ideal observer whose would-be judgments would serve as the basis for conviction, are not arbitrary. That is, because Jim is clear-thinking, unbiased, rational, dispassionate, consistent, and well-informed, the convictions made on the basis of Jim’s would-be judgments would be morally justified.

Bibliography

- Aarts, Henk, Dijksterhuis, Ap [2003] "The Silence of the Library." *Journal of Personality and Social Psychology*, Vol. 84 No. 1, pp. 18-28.
- Aristotle, *De Interpretatione*, chapter 9.
- Ayer, A.J., [1954] "Freedom and Necessity," from *Philosophical Essays*. London: Macmillan, pp. 271-284.
- Balaguer, M., [2009] "Why There are No Good Arguments for any Interesting Version of Determinism," *Synthese* 168:1–21.
- Bargh, John, Chen, Burrows, Mark & Lara, [1996] "The Automaticity of Social Behavior," *Journal of Personality and Social Psychology*, Vol. 71, No. 2.
- Barnes, Elizabeth and Cameron, Ross, [2009] "The Open Future: Bivalence, Determinism, and Ontology," *Philosophical Studies*, Vol. 146, No. 2, pp. 291–309.
- Barnes, Jonathan, [1982] *The Presocratic Philosophers*, rev. ed., London and New York: Routledge.
- Batson, C. D., Sager, K., Garst, E., Kang, M., Rubchinsky, K., & Dawson, K. [1997]. Is empathy-induced helping due to self-other merging? *Journal of Personality and Social Psychology*, 73, 495-509.
- Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1995). "Personal narratives about guilt: role in action control and interpersonal relationships." *Basic and Applied Social Psychology*, 17, 173-198.
- , (1994). "Guilt: An interpersonal approach." *Psychological Bulletin*, 115, 243-267.
- Baumeister, R.F., & Vohs, K.D. [2007] "Self-regulation, Ego Depletion, and Motivation." *Social and Personality Psychology Compass*, 1, 1–14.
- Bishop, Robert C, [2002] "Chaos, Indeterminism, and Free Will," in *The Oxford Handbook on Free Will* edited by Robert Kane. New York: Oxford University Press.
- Bornstein, Brian and Wiener, Richard [2010] *Nebraska Symposium on Motivation: Volume 56: Emotion and the Law: Psychological Perspectives*. New York: Springer.

- Boyd, R., [1972] "Determinism, Laws, and Predictability in Principle," *Philosophy of Science*, Vol. 39, No. 4, pp. 431-450.
- Cashmore, Anthony, [2010] "The Lucretian swerve: The biological basis of human behavior and the criminal justice system," *Proceedings of the National Academy of Sciences*, vol. 107, no. 10, pp. 4499–4504.
- Chisholm, Roderick, [1976] *Person and Object: A Metaphysical Study*, La Salle, IL: Open Court, 1976.
- , [1964] "Human Freedom and the Self," *The Lindley Lectures*, Copyright by the Department of Philosophy, University of Kansas. Reprinted in Pereboom (ed.) 2009.
- Chivers, Tom, [2010] "Neuroscience, Free Will and Determinism: 'I'm Just a Machine'." *The Telegraph*, 12 Oct 2010.
- Clarke, Randolph, [2009] "Dispositions, Abilities to Act, and Free Will: The New Dispositionalism," *Mind*, Vol. 118, Issue: 470, pp. 323-351.
- Cohen, J. [1988] *Statistical power analysis for the behavioral sciences*. Second Edition. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Cotton, Michelle, [2005] "A Foolish Consistency: Keeping Determinism Out of the Criminal Law" *Public Interest Law Journal* Vol. 15, pp. 1-48.
- Dennett, Daniel, [1984] *Elbow Room: The Varieties of Free Will Worth Wanting*, Cambridge, Mass: MIT Press.
- , [1978] "Mechanism and Responsibility," *Brainstorms* (Montgomery, Vermont: Bradford Books.
- Diesmann M., M-O. Gewaltig, and A. Aertsen, [1999] "Stable Propagation of Synchronous Spiking in Cortical Neural Networks." *Nature* 402: 529, 533.
- Dingwall, *et al.* [2003] "Biological Determinism and Symbolic Interaction: Hereditary Streams and Cultural Roads," *Symbolic Interaction*, Vol. 26, No. 4, pp. 631-644.
- Dresser, Rebecca, [1999] "Criminal Responsibility and the 'Genetics Defense'" in *Genetics and Criminality*, (Jeffrey Botkin, William M. McMahon, and Leslie Pickering Francis eds.) Washington DC: American Psychological Association.
- Earman, J., [1986] *A Primer on Determinism*, Dordrecht: Reidel.

- Epstein, Lawrence, [2011] *George Burns: An American Life*. North Carolina: McFarland and Company, Inc.
- Eshleman, Andrew, [2009] "Moral Responsibility," *The Stanford Encyclopedia of Philosophy* (Winter 2009 Edition), Edward N. Zalta (ed.).
- Everett, H., [1973] "The Theory of the Universal Wave Function," in *The Many-Worlds Interpretation of Quantum Mechanics* (B. S. DeWitt & R. N. Graham (Eds.), Princeton, NJ: Princeton University Press.
- Fara, Michael, 2008. "Masked Abilities and Compatibilism," *Mind*, 117 (468): 843–65.
- Feigenson, Neal [2010] "Emotional Influences on Judgments of Legal Blame: How They Happen, Whether They Should, and What to Do About It," in *Nebraska Symposium on Motivation: Volume 56: Emotion and the Law: Psychological Perspectives*, Bornstein and Wiener, eds., pp. 45-96.
- Feinberg, J. & Shafer-Landau, R. [2008] *Reason and Responsibility: Readings in Some Basic Problems of Philosophy*, 13th edition, Belmont: Thomas Wadsworth, p. 408.
- Firth, Roderick. [1952] "Absolutism and the Ideal Observer." *Philosophy and Phenomenological Research*, Vol. 12, No. 3., pp. 317-345.
- Fischer, John Martin, [1994] *The Metaphysics of Free Will*, Oxford: Blackwell Publishers.
- , [1987]. "Responsiveness and Moral Responsibility," In Schoeman (ed.) 1987, 81–106.
- Fischer, John Martin, and Mark Ravizza, [1998]. *Responsibility and Control: An Essay on Moral Responsibility*, Cambridge: Cambridge University Press.
- , [1993] *Perspectives on Moral Responsibility*, Ithaca: Cornell University Press.
- Fodor, J. [1990] *A Theory of Content and other Essays*. MIT Press, p. 196.
- Frankfurt, Harry, [1971] "Freedom of the Will and the Concept of a Person," *Journal of Philosophy*, 68: 5–20.
- , [1969] "Alternative Possibilities and Moral Responsibility," *The Journal of Philosophy*, Vol. 66, No. 23, pp. 829-839.

- Freeman, W. J. and Skarda, C. A. [1987], “How Brains Make Chaos in Order to Make Sense of the World”, *Behavioral and Brain Sciences*, 10: 161–195.
- Gendler, Tamar Szabó, [2008a] “Alief and Belief” *Journal of Philosophy* 105 (10): 634-663.
- , [2008b]. Alief in action (and reaction). *Mind & Language*, 23, 552–585.
- Gere, Cathy, [2011] “Hemispheric Disturbances,” in *The Nation*. December 5, 2011 edition.
- Ginet, Carl, [1996] “In Defense of the Principle of Alternative Possibilities: Why I Don’t Find Frankfurt’s Argument Convincing,” *Philosophical Perspectives*, 10: 403–17.
- Glueck, Sheldon, [1962] *Law And Psychiatry: Cold War Or Entente Cordiale?*
- Habermas, Jürgen, [1990] *Moral Consciousness and Communicative Action*, trans. Christian Lenhardt and Shierry Nicholsen. Cambridge, UK: Polity Press.
- Hardy, Jones, & Gould [1996] *Understanding psychological preparation for sport: theory and practice*. Chichester, UK: Wiley.
- Haug, Matthew [2011] “Explaining the placebo effect: Aliefs, beliefs, and conditioning,” *Philosophical Psychology*, 24:5, 679-698.
- Hemmens, C., Brody, D., Spohn, C. [2013] *Criminal Courts: A Contemporary Perspective, Second Edition*. Thousand Oaks: Sage Publications, Inc.
- Hobbes, Thomas [1997] *Leviathan*, R.E. Flatman & D. Johnston (eds.), New York: W.W. Norton & Co.
- , [1654]. *Of Liberty and Necessity*. Selections reprinted in Raphael, D. D., ed., *British Moralists: 1650–1800*, 2 vols. Oxford: Clarendon Press, 1969. (Vol. 1, 61-70).
- Hodgson, D., [2005] “Quantum Physics, Consciousness, and Free Will,” *The Oxford Handbook of Free Will*, Robert Kane (ed.) Oxford: Oxford University Press.
- Hoefer, Carl, “Causal Determinism,” *The Stanford Encyclopedia of Philosophy* (Spring 2010 Edition), Edward N. Zalta (ed.).

- Hume, David, [1975] *Enquiries concerning Human Understanding and concerning the Principles of Morals*, edited by L. A. Selby-Bigge, third edition, Oxford: Oxford University Press.
- International Social Survey Programme. [1998] Religion II [Data-base]. Retrieved from the German Social Science Infrastructure Services Web site: gesis.org.
- James, William, [1884] “The Dilemma of Determinism” first delivered as an address to Harvard Divinity Students in Lowell Lecture Hall, and published in the *Unitarian Review*, September.
- Johnston, Mark, [1989]: “Dispositional Theories of Value,” Proceedings of the Aristotelian Society, Supplemental Volume 63, 139-174.
- Kane, Robert. [2007]. “Libertarianism.” In *Four Views on Free Will*, by John Martin Fischer, Robert Kane, Derk Pereboom, and Manuel Vargas, 5–43. Oxford: Blackwell.
- Kochen, S., & Specker, E. P., [1967]. The Problem of Hidden Variables in Quantum Mechanics. *Journal of Mathematics and Mechanics*, 17, 59–87.
- Korn, H. & Faure, P [2003] “Is There Chaos in the Brain? Experimental Evidence and Related Models,” *Comptes Rendus Biologies*, Volume 326, pp. 787–840.
- Laplace, P., [1820] *Essai Philosophique sur les Probabilités* forming the introduction to his *Théorie Analytique des Probabilités*, Paris: V Courcier; repr. F.W. Truscott and F.L. Emory (trans.), *A Philosophical Essay on Probabilities*, New York: Dover, 1951.
- Lehrer, Jonah, [2011] “Our Social Networks, Ourselves: Does Free Will Even Exist? Scientists Are Finding that We’re Much More Predictable Than We Think.” *The Wall Street Journal*, U.S. Edition, Saturday, April 23, 2011.
- Lehrer, Keith, [1948] “Cans without Ifs,” *Analysis*, Vol. 29, No. 1, pp. 29-32.
- Lewis, David. [1989]: “Dispositional Theories of Value,” Proceedings of the Aristotelian Society, Supplemental Volume 63, 139-174.
- Libet, B., Gleason, C. A., Wright, E. W., and Pearl, D. K. [1983]. “Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act.” *Brain*, Vol. 106, pp. 623-642.
- Lieberman, J. & Arndt, J. [2000] “Understanding the Limits of Limiting Instructions: Social Psychological Explanations for the Failure of Instructions to Disregard

Pretrial Publicity and Other Inadmissible Evidence” *Psychology, Public Policy, and Law* Vol. 6, No. 3, 677-711.

Lieberman, J. D., Solomon, S., Greenberg, J., & McGregor, H. A. [1999]. A hot new way to measure aggression: Hot sauce allocation. *Aggressive Behavior*, 25, 331-348.

Lief, Michael, Caldwell, Mitchell, and Bycel, Benjamin, [1998] *Ladies and Gentlemen of the Jury: Greatest Closing Arguments in Modern Law*. New York: Scribner.

Luis de Molina, [1988] *On Divine Foreknowledge: Part IV of the Concordia*, tr. Alfred J. Freddoso. Ithaca, New York: Cornell University Press.

Macintyre, A.C., [1957] “Determinism,” *Mind*, New Series, Vol. 66, No 261, pp. 28-41.

Maner, J. K., Luce, C. L., Neuberg, S. L., Cialdini, R. B., Brown, S., & Sagarin, B. J. [2002]. The effects of perspective taking on helping: Still no evidence for altruism. *Personality and Social Psychology Bulletin*, 28, 1601-1610.

Mansnerus, Laura and Hanley, Robert, [2001] “Death Sentence Overturned For Killer of Deliverymen,” *The New York Times*. June 8th. Retrieved from <http://www.nytimes.com>.

Maudlin, T., [2007] *The Metaphysics Within Physics*. Oxford: Oxford University Press.

McKenna, Michael, [2011] “Contemporary Compatibilism: Mesh Theories and Reasons-Responsive Theories” in *The Oxford Handbook of Free Will, Second Edition*, Robert Kane (ed.). Oxford: Oxford University Press.

—, [2009] “Compatibilism”, *The Stanford Encyclopedia of Philosophy (Winter 2009 Edition)*, Edward N. Zalta (ed.).

—, [2000]. “Assessing Reasons-Responsive Compatibilism,” *International Journal of Philosophical Studies*, 8 (1): 89–114.

McKenna, Michael and Russell, Paul (eds.), [2008] *Free Will and Reactive Attitudes: Perspectives on P. F. Strawson's Freedom and Resentment*, Abingdon: Ashgate Publishing Group.

Meir Dan-Cohen, [1983] “Actus Reus.” In S.H. Kadish, ed., *Encyclopedia of Crime and Justice* (New York: Free Press).

Mele, Alfred, [2006]. *Free Will and Luck*, New York: Oxford University Press.

- , [2000]. “Reactive Attitudes, Reactivity, and Omissions,” *Philosophy and Phenomenological Research*, 61: 447–452.
- Mele, Alfred, and David Robb, [1998] “Rescuing Frankfurt-Style Cases,” *Philosophical Review*, 107: 97–112.
- Menninger, Karl, [1966] *The Crime of Punishment*.
- Moore, G.E., [1912] *Ethics*, Home University Library, ch. 6.
- Moore, Michael, [2010] “Causation and the Excuses,” in *Placing Blame: A Theory of the Criminal Law*. Oxford: Oxford University Press.
- Mulligan, K., Simmons, P., and Smith, B., [1984] “Truth-Makers,” *Philosophy and Phenomenological Research*, Vol. 45, No. 3, pp. 287–321.
- Nadelhoffer & Matveeva. [2009] “Positive Illusions, Perceived Control and the Free Will Debate,” *Mind & Language*, Vol. 24, No. 5, pp. 495–522.
- Naylor, Marjory, [1984] “Frankfurt on the Principle of Alternate Possibilities,” *Philosophical Studies*, 46: 249–58.
- Paulhus, D.L., & Margesson, A. [1994] Free Will and Determinism (FAD) scale. Unpublished manuscript, University of British Columbia, Vancouver, British Columbia, Canada.
- Payment, Simone, [2004] *The Trial of Leopold and Loeb: A Primary Source Account*. New York: The Rosen Publishing Group, Inc.
- Peet, Richard, [1985] “The Social Origins of Environmental Determinism,” *Annals of the Association of American Geographers*, Vol. 75, No. 3, pp. 309–333.
- Penrose, R. [1997] in *The Large, the Small and the Human Mind*, (ed. by M. Longair). Cambridge, England: Cambridge University Press.
- , [1989] *The Emperor’s New Mind*. New York: Oxford University Press.
- Pereboom, Derk, [2001]. *Living Without Free Will*, Cambridge: Cambridge University Press.
- , [2009] *Free Will 2nd edition* (Pereboom ed.) Cambridge: Hackett Publishing Company, Inc.
- , [2007]. “Hard Incompatibilism.” In *Four Views on Free Will*, by John Martin Fischer, Robert Kane, Derk Pereboom, and Manuel Vargas, 85–125. Oxford: Blackwell.

- , [2006] “Why We Have No Free Will and Can Live Without It,” reprinted in *Reason and Responsibility: Readings in Some Basic Problems of Philosophy*, 13th ed. (Feinberg, J. & Shafer-Landau, R. ed.) Belmont: Thomas Wadsworth, p. 464-477.
- Pettit P., [1991] “Realism and Response-Dependence,” *Mind* 100, no. 399, 587-626.
- Plato. [2006], *Republic*, translated and with intro by R.E. Allen, New Haven: Yale University Press.
- Rigoni, D., Kuhn, S., et al. [2011] “Psychological Inducing Disbelief in Free Will Alters Brain Correlates of Preconscious Motor Preparation: The Brain Minds Whether We Believe in Free Will or Not,” *Psychological Science* 22: 613-618.
- Robinson, M. D., & Clore, G. L. [2002]. Episodic and semantic knowledge in emotional self-report: Evidence for two judgment processes. *Journal of Personality and Social Psychology*, 83, 198-215.
- Ross, Craig. [2011]. “Dennett on Free Will.” *Metaphysica* 12, no. 2: 137-149.
- Rostbøll, Christian. [2010] “Freedom as Satisfaction? A Critique of Frankfurt’s Hierarchical Theory of Freedom” *SATS: Northern European Journal of Philosophy*. Volume 5, Issue 1: 131–146.
- Rotter, J.B. [1954] *Social learning and clinical psychology*. NY: Prentice-Hall.
- Rozin, P., Millman, L., & Nemeroff, C. [1986]. Operation of the laws of sympathetic magic in disgust and other domains. *Journal of Personality and Social Psychology*, 50, 703–712.
- Russell, Paul, [1995] *Freedom and Moral Sentiment*, New York: Oxford University Press.
- Scanlon, T.M., [1988] “The Significance of Choice,” in Sterling M. McMurrin, ed., *The Tanner Lectures on Human Values*, vol. 8 (Salt Lake City: University of Utah Press), pp. 149-216, 165-166, and 169-170.
- Schoeman, Ferdinand (ed.), 1987. *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, Cambridge: Cambridge University Press.
- Slobogin, Christopher, [2005] “The Civilization of the Criminal Law,” *Vanderbilt Law Review*, Vol. 58:1:121.
- Slote, Michael, [1980] “Understanding Free Will,” *Journal of Philosophy*, 77: 136–51.

- Smilansky, S. [2001]. "Free Will: From Nature to Illusion." *Proceedings of The Aristotelian Society*, 101, 71–95.
- Snow, C.P. [1983], in John Halperin *C.P. Snow: An Oral Biography* (Harvester, Brighton).
- Sobel, Jordan H. [1998] *Puzzles for the Will: Fatalism, Newcomb, and Samarra, Determinism and Omnipotence*. Toronto: University of Toronto Press Inc.
- Southwood, Nicholas, [2010] *Contractualism and the Foundations of Morality*. Oxford: Oxford University Press.
- Stillman, Tyler, et al. [2010]. "Personal Philosophy and Personnel Achievement: Belief in Free Will Predicts Better Job Performance," *Social Psychological and Personality Science*, January 2010, pp. 43-50.
- Stillman, T.F & Baumeister, R.F. [2010] "Guilty, Free, and Wise: Determinism and Psychopathy Diminish Learning from Negative Emotions," *Journal of Experimental Social Psychology*, Vol. 46, pp. 951–960.
- Strawson, P. F., [1962] "Freedom and Resentment," *Proceedings of the British Academy*, 48: 187–211.
- Stump, Eleonore, [1996]. "Persons: Identification and Freedom," *Philosophical Topics*, 24: 183–214.
- Taylor, C.C.W., [1999] *The Atomists: Leucippus and Democritus. Fragments, A Text and Translation with Commentary*, Toronto: University of Toronto Press.
- Taylor, Richard, [1962] "Fatalism," *The Philosophical Review*, Vol. 71, No. 1, pp. 56-66.
- Tegmark, M. [2000] "Importance of Quantum Decoherence in Brain Processes," *Physical Review, Series E*. Vol 61; Part 4; Part B, pp. 4194-4206.
- Theodorakis & Chroni [2001] "Self-talk in a basketball-shooting task." *Perceptual and Motor Skills*. Feb: 92(1): 309-15.
- Treadwell, Kimberly R. H.; Kendall, Philip C. [1996] "Self-talk in youth with anxiety disorders: States of mind, content specificity, and treatment outcome." *Journal of Consulting and Clinical Psychology*, Vol 64(5), Oct, 941-950.
- Twenge, J.M., Zhang, L., & Im, C. (2004). It's beyond my control: A cross-temporal meta-analysis of increasing externality in locus of control, 1960-2002. *Personality and Social Psychology Review*, 8, 308–319.

- Van Inwagen, P [1999] "Moral Responsibility, Determinism, and The Ability to Do Otherwise," *The Journal of Ethics* Vol. 3, No. 4, pp. 341-350.
- , [1983] *An Essay on Free Will*. Oxford: Claredon Press.
- , [1975] "The Incompatibility of Free Will and Determinism," *Philosophical Studies* Vol. 27, pp. 185-199.
- , [1974] "A Formal Approach to the Problem of Free Will and Determinism." *Theoria* 40: 9-22.
- Van Raalte, Judy L.; Brewer, Britton W et al. [1995] "Cork! The effects of positive and negative self-talk on dart throwing performance." *Journal of Sport Behavior*, Vol 18(1), Mar 1995, 50-57.
- , [1994] "The relationship between observable self-talk and competitive junior tennis players' match performances. *Journal of Sport & Exercise Psychology*, Vol 16(4), Dec 1994, 400-415.
- Verme, Paolo. [2009] "Happiness, Freedom and Control," *Journal of Economic Behavior & Organization* Vol. 71, pp. 146-161.
- Vihvelin, Kadri, [2004] "Free Will Demystified: A Dispositional Account," *Philosophical Topics*, 32: 427–50.
- Vohs, Kathleen & Baumeister, Roy. [2009] "Addiction and Free Will," *Addiction Research and Theory*, Vol. 17:3, pp. 231–235.
- Vohs, Kathleen & Schooler. [2008] "The Value of Believing in Free Will: Encouraging a Belief in Determinism Increases Cheating," *Psychological Science* Vol. 19: 49, pp. 49-54.
- Wallace, Jay, [1994] *Responsibility and the Moral Sentiments*. Cambridge, Massachusetts: Harvard University Press.
- Walton, Kendall [1978] "Fearing Fiction," *The Journal of Philosophy*, Vol. 75, No.8.
- Warfield, Ted A. [1997] "Divine Foreknowledge and Human Freedom Are Compatible," *Noûs*, Vol. 31, No. 1, pp. 80-86.
- Watson, Gary, [1988] "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme," in McKenna, Michael and Russell, Paul (eds.), [2008] *Free Will and Reactive Attitudes: Perspectives on P. F. Strawson's Freedom and Resentment*, Abingdon: Ashgate Publishing Group.

- , [1975] “Free Agency,” *Journal of Philosophy*, 72: 205–20.
- West, S. G., & Brown, T. J. [1975]. Physical attractiveness, the severity of emergency and helping: A field experiment and inter- personal simulation. *Journal of Experimental Social Psychology*, 11, 531-538.
- Widerker, David, [1995] “Libertarianism and Frankfurt’s Attack or the Principle of Alternative Possibilities,” *Philosophical Review*, 104: 247-61.
- Wiggins, D. [1987] *Needs, Values, Truth*. Cambridge, MA: Blackwell Publishers, Inc.
- Wolf, Susan, 1990. *Freedom within Reason*, Oxford: Oxford University Press.
- Wooton, Barbara, [1963] *Crime And The Criminal Law* 32-84.
- Wright, Robert, [1994] *The Moral Animal*, New York: Pantheon Books.
- Zagzebski, Linda, [1985] “Divine Foreknowledge and Human Free Will,” *Religious Studies*, Vol. 21, No. 3, pp. 279-298.