UNIVERSITY OF CALIFORNIA
RIVERSIDE

Freedom and Explanation: A Defense of the Fixity of the Independent

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

Doctor of Philosophy

in

Philosophy

by

Andrew Robert Law

June 2020

Dissertation Committee:
Dr. Michael Nelson, Chairperson
Dr. John Martin Fischer
Dr. Erich Reck
Dr. Luca Ferrero
Dr. Adam Harmer
Dr. Carolina Sartorio

The Dissertation of Andrew Robert Law is approved:

____

____

____

____

____

____

Committee Chairperson

Acknowledgments

I suppose I wrote (most of) the words of this dissertation, but it would be vanity to think this dissertation belongs to me alone, as there are far too many individuals who have contributed to this work in one way or another. All of my committee members have offered valuable insights on previous drafts. A special thank you to Dr. Michael Nelson and Dr. John Fischer for discussing the ideas contained within many, many times over the years. I am also extremely grateful to Dr. Carolina Sartorio for agreeing to serve on my committee despite not being affiliated with UCR, and for the excellent comments she has given on several chapters.

No one has read more drafts or given more feedback than Dr. Taylor Cyr and Jonah Nagashima, and I will always feel indebted to them for encouraging me to join their little "Free Will Reading Group." I have no doubt that this dissertation wouldn't exist if they hadn't invited me (multiple times) to participate. And an additional thank you to Dr. Taylor Cyr who co-authored a publication with me that is embedded in chapter 3. The text of that chapter is, in part, a reprint of the material as it appears in "Freedom, Foreknowledge, and Dependence: A Dialectical Intervention," forthcoming in *American Philosophical Quarterly*. Additionally, the text of chapters 1 and 2 are, in part, reprints of the material as it appears in "From the Fixity of the Past to the Fixity of the Independent," forthcoming in *Philosophical Studies* as well as "The Dependence Response and Explanatory Loops," forthcoming in *Faith and Philosophy*.

Finally, I thank my family for all of the support, both emotional and financial, that they have given me while crafting this dissertation. You have all offered encouragement and even shown interest in my work over the years, something no philosopher takes for granted. And above all, thank you to my wife, Meghan, who moved to a place too bright and too hot, worked a job too long for too little, all so I might one day have a few extra letters next to my name. Thank you.

Dedicated to my family by blood, my family by marriage, and my family by friendship: The Laws, In-Laws, and Out-Laws

ABSTRACT OF THE DISSERTATION

Freedom and Explanation: A Defense of the Fixity of the Independent

by

Andrew Robert Law

Doctor of Philosophy, Graduate Program in Philosophy
University of California, Riverside, June 2020
Dr. Michael Nelson, Chairperson

In the first three chapters, a novel principle of agency dubbed the principle of the "Fixity of the Independent" (or "FI") is developed and defended. FI holds, roughly, that if a fact is in no way explained by an agent's current behavior, then the fact is currently beyond the agent's control or "fixed." In the first and second chapter, it is argued that FI is the best explanation of more familiar principles of agency, such as the principle that the past is currently beyond every agent's control. According to FI, the past is currently beyond every agent's control precisely because the past is in no way explained by any agent's current behavior. In the third chapter, it is argued that FI delivers unique results in various debates about freedom: while FI implies that causal determinism would threaten our freedom, it suggests that neither a fully determinate future nor the existence of an essentially all-knowing being would. In the fourth chapter, the notion of explanation is

utilized for a slightly different purpose, namely, to answer a pressing challenge to a popular view of freedom, a view strongly supported by FI. The appendix deals with a complication brought out in the third chapter.

Table of Contents

Chapter 1: From the Fixity of the Past to the Fixity of the Independent

Chapter 2: Objections to the Fixity of the Independent

Chapter 3: Implications of the Fixity of the Independent

Chapter 4: Explanation and the Problem of Enhanced Control

List of Figures

Preface

We have some control over our lives, but not complete control. Maybe there was nothing you could have done to get that promotion at work; perhaps the dream car was never meant to be; certainly, the result of the last presidential election wasn't (entirely) up to you. But eating the donut for breakfast instead of the apple? Or wearing a red shirt today rather than a blue one? You had control over those things. Or so it seems.

Reflections like these, however mundane and commonsensical, quickly give rise to difficult questions. When, exactly, is something under our control? Can we really be so sure that we have *any* control over our lives? What are we to make of the various challenges to control that have been given over the centuries, challenges from areas as disparate as logic and biology, religion and physics? When philosophers talk about "the problem(s) of free will," it is questions like these (among others) that they have in mind.

The goal of this dissertation is to make some headway in answering such questions. In the first two chapters, I articulate and defend a particular principle of freedom, what I call the "Principle of the Fixity of the Independent" (or "FI"). Roughly, FI holds that if a part of the world is in no way explained by our current behavior, then that part of the world is beyond our control; and so if our performing a certain action now would require that part of the world to be different, then we are not free to perform the action in question. Although novel, this principle undergirds many other intuitive ideas and principles of freedom, or so I argue. In the third chapter, I then apply this principle to three of the most prominent challenges to freedom: the argument from future contingents, the argument from divine foreknowledge, and the argument from causal determinism. I argue that, out of the bunch, only causal determinism would pose a threat to our freedom, at least if FI is correct. These three chapters constitute the bulk of the dissertation, helping us get clear on when a part of the world is under our control and what kinds of discoveries would undermine our freedom.

In the fourth chapter, I discuss a challenge not aimed directly at FI, but rather at a family of views closely associated with it. The challenge is often called “The Problem of Enhanced Control” and it targets any view which claims that, although causal determinism would undermine our freedom (as FI implies), we are nevertheless free to do otherwise on certain occasions. I present a new solution to this problem, one where the notion of explanation again plays a central role, and argue that it is superior to extant solutions.

The dissertation concludes with an appendix which deals with some of the complications glossed over in the third chapter. So-called “classical compatibilists” hold that causal determinism would not undermine our freedom to do otherwise, contrary to what FI implies. What should such authors make of the arguments presented for FI? The appendix engages this question by examining the work of perhaps the most influential classical compatibilist, David Lewis. It is argued that, while Lewis’s version of classical compatibilism may be immune to the arguments for FI, there are significant problems for Lewis's version. Moreover, once we amend Lewis's version so as to overcome these problems, we end up with a version of classical compatibilism that is susceptible to the arguments for FI.

Finally, let me note what the dissertation does *not* attempt to do. Nowhere does it offer a fully developed account of the notion of explanation. I do make some substantive assumptions about the nature of explanation throughout, noted primarily in the first and fourth chapters, but even these are mostly for expository purposes. This lack of a full account might seem a bit strange (to say the least) since the notion of explanation plays such a pivotal role throughout the entire dissertation. But in my view, a full account is not needed. If anything, it would be unfortunate if the central points of the dissertation hinged on the relatively minute differences between particular accounts. Instead, I take my claims about various explanatory connections (or lack thereof) to be both fairly intuitive and compatible with most accounts of explanation.

For the insistent reader though, let me note three features of the notion of explanation that seem as close to non-negotiable as they come. First, the explanatory relation takes facts as relata, where a fact is something like a true proposition. (Although sometimes for ease of expression, I'll allow events to serve as relata.) How many facts does it take? I'll typically invoke a non-contrastive view of explanation where only two facts are required – the one being explained and the one doing the explaining – but I believe everything I say is compatible with contrastive views where three or four facts are required instead. Second, the explanatory relation is asymmetric, although even here there is some wiggle-room. If it is possible for there to be explanatory loops (as I think) and explanation is transitive (as I'm open to), then the asymmetry of explanation will be violated. But we can get around this difficulty by claiming that explanation is at least *directly* asymmetric, where that means if fact *A* explains fact *B directly*, then *B* does not explain *A directly*; if *B* does explain *A*, it is only *ancestrally*. Third, and perhaps most substantively, if *A* explains *B*, then *A* answers a "why" question about *B*, at least in some possible context. The fact that I threw a rock at the window explains why the window shattered, or the fact that there is oxygen in the room explains why the match lit when struck, etc., because in each case the former fact tells us *why* the latter fact obtained, at least in some possible context. This third feature is where things can get very deep very fast, as issues about the pragmatics of explanation (and pragmatism more generally) are just around the corner. I'll be assuming a more "realist" view of explanation: that the explanatory relation is an objective feature of the world (or at least tracks such features), although I think more "anti-realist" views aren't necessarily antithetical to the claims made in the dissertation.

Admittedly, many of these remarks are promissory. At the very least then, I would encourage the reader to think of the dissertation as giving us a new desideratum for an account of explanation. Given the incredible appeal of principles like FI, an account of explanation that further justified the claims made throughout the dissertation would have quite a bit going for it.

The other issue that the dissertation does not attempt to address is how to best distinguish between the myriad notions involved in free agency. The notion that philosophers throughout the ages have (arguably) focused on the most, and the one that is the centerpiece of this dissertation, is the *freedom to do otherwise*. But there are several other notions in the area, with sharp disagreement over how such notions are connected. The most famous example is the relation between the freedom to do otherwise and moral responsibility. Up until about fifty years ago, almost everyone assumed that being morally responsible for an action required being free to do otherwise than that action. But then Harry Frankfurt came along, along with a slew of others shortly thereafter (including prominent members on this dissertation committee), and argued forcefully that this was mistaken: that one could *freely* perform an action, thereby possibly being morally responsible for it, even if one wasn't *free to do otherwise*. Something similar has happened with the relations between other agential notions as well, such as the notions of autonomy, identification, and even agency itself.

Working through this tangled web would be a dissertation of its own, so I'll instead try to remain agnostic about these issues as best I can. Even if the freedom to do otherwise isn't required for moral responsibility, say, it still seems to be something we value – I presume that most of us would be disappointed to learn that this freedom was illusory, regardless of whether that discovery changed the ways in which we held each other accountable. So it still seems worthwhile to get clearer on this particular notion. Again, sometimes for ease of expression, I'll use locutions such as “freedom” and “freely,” but unless explicitly stated otherwise, these should be taken as shorthand for “free to do otherwise.” Nevertheless, the connections between various agential notions cannot be completely avoided. They appear here and there throughout the dissertation and come to a head in the fourth chapter, where I explore ways of understanding and unifying various agential notions, including the notions of freedom *simpliciter* and authorship. But the reader should approach the dissertation, first and foremost, as a project intended to help us better understand the freedom to do otherwise and what to make of the various threats to that particular kind of freedom.

Chapter 1: From the Fixity of the Past to the Fixity of the Independent

Introduction

The Principle of the Fixity of the Past (FP henceforth) holds, roughly, that if an agent's performing a certain action requires that the past be different, then the agent cannot perform the action in question. For instance, suppose that a student, Sophia, now wishes to take a seminar from David Lewis. (Sophia is impressed by Lewis's work but unaware of his premature death.) In order for Sophia to take a seminar from Lewis, the past would have to be different – at the very least, her taking a seminar now would require Lewis to have survived 2001. Thus, FP implies that Sophia cannot take a seminar from Lewis.

While FP is controversial, many admit that it (or some similar principle) is quite intuitive at first glance. And given its intuitive appeal, it has come to play a central role in many debates over the freedom to do otherwise. Prominent examples include the debates over freedom and future contingents, freedom and divine foreknowledge, and freedom and determinism.[1]

Recently though, several authors have argued (or at least implied) that FP ought to be abandoned in favor of an alternative principle, what I will be calling

[1] See Aristotle (ch. IX), Pike (1965), and van Inwagen (1983), respectively.

the "Principle of the Fixity of the Independent" ("FI" henceforth). According to FI, it is not the past per se that is beyond our control or "fixed." Rather, it is those facts that are wholly "independent" of the agent's present behavior which are fixed. For instance, in the case of Sophia, Lewis's death is not beyond her control or "fixed" for her because it is *past*; rather, it is beyond her control or "fixed" for her because Lewis's death is not "dependent," in the relevant way, on anything she presently does.

Now, it may generally be true that past facts are wholly independent of any particular agent's present behavior, but it is not necessarily true (or so it is claimed). And if there is a case where a past fact depends on an agent's present behavior, FI does not imply (*contra* FP) that the fact in question is fixed for the agent. With this distinctive implication, certain facts that might seem to threaten freedom under FP do not seem so threatening under FI.

For example, consider the case of divine foreknowledge and freedom. Suppose that at time t_1, God knows (and hence believes) that agent *S* will perform action *X* at time t_2 (where t_1 is earlier than t_2). As many have argued, FP seems to imply that *S* is not free to do otherwise than *X* at t_2. Given God's essential omniscience, *S*'s doing otherwise than *X* at t_2 seems to require that the past relative to t_2 be different – it seems to require that, at t_1, God not believe that *S* will perform *X* at t_2. So FP seems to imply that divine foreknowledge is incompatible with the freedom to do otherwise.

But now consider what FI has to say about divine foreknowledge and freedom. Arguably, the fact that, at t_1, God believes that *S* will perform action *X* at t_2 is "dependent" on the fact that *S* performs *X* at t_2. For instance, perhaps God only has the relevant belief at t_1 because all times, including t_2, are "present" or "available" to God. But if God's belief at t_1 is dependent on *S*'s behavior at t_2, then FI does not imply that God's belief at t_1 is fixed for *S* at t_2. So FI does not seem to imply that divine foreknowledge is incompatible with the freedom to do otherwise.[2]

Arguably then, there are some facts which are a threat to freedom under FP but not under FI. It also goes the other way: there seem to be some facts that are a threat to freedom under FI but not under FP. For instance, suppose there are *future* facts that are completely independent of our current behavior. FI, but not FP, implies that such facts are fixed for us currently.[3] (Indeed, this point will play a role in some of the arguments to follow.) So while there is a good amount of overlap between FP and FI, they are importantly distinct.

In this chapter, I offer four arguments for why those sympathetic to FP ought to abandon it in favor of FI. In particular, I will be arguing that FI is the more general principle behind FP, and that FP is at best a derivative and contingent principle. The first two arguments are nascent in the literature,[4] while the last two

[2] For defenses of something like FI and its application to divine foreknowledge, see Merricks (2009; 2011), McCall (2011), Westphal (2011), and Swenson (2016). For dissent, see Fischer & Todd (2011), Todd & Fischer (2013), and Fischer & Tognazzini (2014).

[3] Thank you to Carolina Sartorio for helping me see this.

[4] See Merricks (2009; 2011).

are wholly original to this chapter. And while I don't think any argument by itself constitutes a conclusive argument for FI (or against FP), I do think that, when taken collectively, they give us strong reason to prefer FI over FP.

Before getting to the arguments, though, it is necessary to frame FP, FI, and the dialectic concerning them.

Framing FP and FI

There is a good deal of controversy over FP and how it should be formulated.[5] Instead of delving into these issues too much, I will use an extremely weak formulation:

> FP: Agent *S* can perform action *X* at time *t* (in world *w*) only if there is a world, *w'*, such that *w'* has the same past as *w* up until *t* and *S* performs *X* at *t*.[6]

As noted, this formulation of FP is very weak: it only requires that the past be *consistent* with the action in question. For our purposes, this is a good thing. If it can be shown that this formulation of FP ought to be abandoned in favor of FI, that is a stronger result.

Now in framing FI, it is crucial to get clear on what notion of "dependence" is at stake. While all authors agree that the notion of "dependence" in question is

[5] For the debate over FP, see Fischer (1994, ch. 4), Holliday (2012), and Dorr (2016).
[6] This formulation is slightly amended from Fischer (2011).

stronger than mere counterfactual dependence, agreement ends there. Some authors, like Jonathan Westphal (2011), use the notion of "supervenience"; others, like Trenton Merricks (2009; 2011), use a simple "because"; and the list of variations goes on. In formulating FI, I will follow a particular proposal given by Philip Swenson (2016), and for two reasons. First, his account of dependence is one of the most developed accounts (in this context). Second, his notion of dependence seems to provide at least a necessary condition on the other accounts of dependence, e.g., one fact is dependent on another in Merricks's sense only if it is also dependent in Swenson's sense. Insofar as Swenson's notion provides a necessary condition, it provides a "common thread" through these other accounts.

Swenson suggests we understand the kind of dependence in FI as *explanatory dependence*. Roughly, a fact, F_1, is explanatorily *dependent* on another fact, F_2, just in case F_2 at least partly explains F_1; F_1 is thus explanatorily *independent* of F_2 just in case F_2 does not even partly explain F_1.[7] The notion of explanation here is meant to be very broad: it is meant to include logical, conceptual, metaphysical, and even nomic/causal explanation. For example, the fact that the ball is colored is explanatorily dependent on the fact that it is red; the fact that the singleton {Socrates} exists is explanatorily dependent on the fact that Socrates exists; the fact that the window shattered is explanatorily dependent on the fact that Suzy threw a rock at it.[8]

[7] For expository purposes, I will sometimes talk as if the "explanatory dependence" relation takes events as relata, not facts. Nothing substantive hangs on this issue.

[8] See Swenson (2016, pp. 660-61) for discussion of similar examples.

With this notion of explanatory dependence in mind, I will understand FI as follows:

> FI: Agent *S* can perform action *X* at time *t* (in world *w*) only if there is a world, *w'*, such that all of the facts that are distinct from and explanatorily independent of the facts constituting *S*'s behavior at time *t* hold and *S* performs *X* at *t*.[9]

Intuitively, FI claims that any fact which is in no way explained by the agent's behavior is beyond the agent's control or "fixed" for the agent. Hence, if performing a certain action would require such a fact to be different, then the agent cannot perform the action in question. Consider the case of Sophia again. Plausibly, the fact that Lewis died in 2001 is in no way explained by Sophia's current behavior. So, FI implies that his death is fixed for Sophia now. And given that every world where Sophia now takes a seminar from Lewis is a world where Lewis survives 2001, FI implies that Sophia cannot take a seminar from Lewis.[10]

While I will proceed with these more precise understandings of FP and FI, I do not think the success of my arguments hangs (too much) on any particular explication. But before moving on to these arguments, it is important to first frame the dialectic concerning FP and FI. When defending FP, authors typically offer

[9] This formulation is inspired by, albeit distinct from, Swenson's formulation (2016). Swenson restricts his formulation to the independent *past*.

[10] It is dubious that every world where Sophia now takes a seminar from Lewis is a world where Lewis survives 2001. Just consider worlds where Lewis is miraculously resurrected. But such worlds seem to be ones where the laws of physics and biology are quite different than our own. Given that the laws of physics and biology are also explanatorily independent of Sophia's current behavior, such worlds can be ruled out too. Thank you to Eric Funkhouser for noting this.

paradigm cases of past facts that seem to be beyond our control or fixed. For instance, consider the fact that John F. Kennedy was killed in 1963. Intuitively, this fact is beyond our control – there is nothing any of us can now do about Kennedy's assassination. These authors then claim that FP captures our intuitions in these kinds of cases.

Note, though, that FI also captures our intuitions in such paradigm cases. Plausibly, Kennedy's assassination is in no way explained by anything that any of us are currently doing. Hence, FI also implies that his assassination is fixed for us now. If so, then such paradigm cases do not obviously support FP over FI. Indeed, given that the past is (generally) not explained by any agent's current behavior, it's hard to think of an uncontroversial case that would clearly support FP without also supporting FI. Given that these paradigm cases are usually the central pieces of support for FP, it seems then that FP and FI are initially on even ground: that we should start off more-or-less agnostic as to whether FP or FI is to be preferred. With that in mind, I now give four arguments for why FI is to be preferred, and that FP is at best *derivative* of FI.

The Asymmetry Between Past and Future

While it is quite intuitive that the past is beyond our control or fixed, it is quite *unintuitive* that the future is beyond our control or fixed. On the contrary, most of us think of the future as being (somewhat) under our control. But what explains the difference? Why is the past beyond our control but not the future?

Here is part of an answer: the past is beyond our control or “fixed” because it is beyond our “causal reach.” No effect precedes its cause. Hence, nothing we do now can have any kind of causal effect on the past. But the future is plainly within our causal reach – many of the things we do now could have a causal effect on the future.[11]

This is only a partial answer, though. It’s clear that in order for this answer to be relevant, we must make something like the following presupposition: if a part of the world is beyond an agent’s causal reach, then that part of the world is fixed for the agent; otherwise, it isn’t (necessarily) fixed for the agent. But that presupposition is right in line with FI. After all, if some event is beyond an agent’s causal reach, then the event is in no way causally explained by the agent’s behavior. Assuming that the event is not explained by the agent’s behavior in some non-causal way, which seems plausible in the typical case, it follows that the event is *explanatorily independent* of the agent’s behavior. So, FI seems to imply that, in the typical case, if a part of the world is beyond an agent’s causal reach, then that part of the world is fixed for the agent.[12]

[11] See Mellor (1998, ch. 10) for an especially interesting discussion of this claim.

[12] FI implies that this principle only holds in the typical case because there may be instances where an event is beyond the causal reach of the agent but not beyond her “non-causal” reach. One such instance has to do with so-called “soft facts” about the past, which are discussed in the next section.

Thus FI, in conjunction with the claim that effects do not precede their causes, has a straightforward and powerful explanation for the asymmetry between the past and future: generally, any part of the world that is beyond our causal reach is beyond our control, and the past, but not necessarily the future, is beyond our causal reach. It's important to notice that none of this implies that FP is false. On the contrary, the asymmetry between past and future presupposes that something like FP is (generally) true. Nonetheless, there are two important and related implications to be drawn. First, it implies that FP, if true, is only *contingent*. Consider worlds where effects routinely precede their causes, where there is rampant "backwards causation." In such worlds, FI does not deliver the same results as FP. For agents living in worlds with rampant backwards causation, the past is often within their causal reach and, hence, partly explained by their current behavior. FI does not imply that the past is fixed for such agents.[13]

Even some of the most ardent defenders of FP recognize this point. For instance, Fischer & Todd (2011, p. 105 n14) say "…if we can causally affect the past or directly change it, this would certainly call (FP) into question." They even go so far as to say that the absence of backwards causation is a "prerequisite" of FP (p. 104). But this admission, that FP presupposes an absence of backwards causation, leads to the second implication of the argument above: that FP, even if true, is only a derivative principle, and that FI (or something similar) is the deeper, more

[13] See Sartorio (2015) for a helpful discussion of the past being beyond our causal reach and the contingency of the fixity of the past.

general principle. But if FI is the deeper and more general principle, then in cases where FP and FI deliver distinct results, FI ought to be preferred.

Perhaps the jump to FI here is too quick, though. In particular, let's explore some other ways that one could justify the asymmetry between past and future without invoking FI. I can think of three plausible suggestions: (i) appeal to the so-called "growing-block" theory of time; (ii) appeal to the past being "over and done with"; and (iii) appeal to some principle substantially weaker than FI. Let's take each suggestion in turn.

The growing-block theory of time holds, roughly, that the past and present are real but the future is not. The universe "grows" as time moves on. Many authors endorse the growing-block theory precisely because the past seems fixed while the future does not. According to these authors, the fact that the past exists, but the future does not, explains this asymmetry.[14] If this explanation is successful, then it would seem to vindicate FP without a commitment to anything like FI.

Even if the growing-block theory gives a satisfactory explanation of the asymmetry between past and future (which is controversial), there are at least two reasons why FI is to be preferred. First, endorsing the growing-block theory is costly. Plausibly, endorsing the growing-block theory also requires endorsing the *A*-theory of time and all of the troubles that besiege it.[15] But there are also problems

[14] For a thorough defense of the growing-block theory, see Tooley (1997).

[15] There are many objections to the *A*-theory and its different versions. See Mellor (1998) and Sider (2001).

particular to the growing-block theory. For instance, consider the problem of future contingents: how can a proposition about the future be (contingently) true if the future doesn't exist? Or consider what is sometimes called the "epistemic" problem: if the past is just as real as the present, how do we know we are living in the present, not the past? How do we know that we occupy the "cutting edge" of reality, not some previously added part?[16]

This is not the place to delve into these complex issues, and I certainly do not mean to suggest that they are insoluble. Rather, my point is much simpler, namely, that embracing the growing-block theory to explain the asymmetry between past and future is (potentially) costly, at least when compared to the alternative of embracing FI and the innocuous claim that effects do not precede their causes. Since FI seems compatible with most any view in the philosophy of time,[17] all of these (potential) problems can be avoided. FI seems to be the more ecumenical explanation. Surely, that is a point in favor of FI.

But there is a second reason to prefer FI over the growing-block theory. Even if the growing-block theory successfully explains the asymmetry of control over past and future, there is another asymmetry involving control that it fails to explain. While we presumably have control over some parts of the future, there seem to be parts of the future that we plainly *do not* have control over. For instance,

[16] See Briggs & Forbes (2012) for a response to the problem of future contingents. See Merricks (2006) and Forrest (2006) for discussion of the epistemic objection.

[17] Some might argue that FI delivers the right results only if views like *presentism* in the philosophy of time are false. See Finch & Rea (2008) for an argument along these lines. See chapter 2 for discussion of this argument.

while we seem to have some control over what happens on the surface of the Earth in the next ten minutes, we do not seem to have any control over what happens on the surface of Pluto in the next ten minutes.

What explains this difference? Why are only some parts of the future under our control? Here's a familiar suggestion: generally, in order to have control over a part of the world, it needs to be within our causal reach. Given that much of the future, such as what happens on Pluto in the next ten minutes, is beyond our causal reach, it follows that much of the future is beyond our control. But, again, this suggestion is right in line with FI.

It's not clear how the growing-block theory itself could explain this second asymmetry. After all, according to the growing-block theory, *none* of the future exists – all events in the future are ontologically "on a par." It's unclear, then, how the growing-block theory could even start to give an explanation for why some, but not all, of the future is under our control.

To be clear, I am not saying that the growing-block theorist is prohibited from offering an explanation for this second asymmetry. On the contrary, I encourage the growing-block theorist to endorse FI! What I am saying is that the growing-block theory *itself* cannot explain this asymmetry. In contrast, FI can explain both this asymmetry and the asymmetry between past and future: it explains why we lack control over what *happened* on Pluto ten minutes ago as well as what *will happen* on Pluto ten minutes from now. FI provides a uniform explanation for these asymmetries and, thus, has more explanatory power. That is a strong reason to prefer FI over the growing-block theory.

Let's move on to the second mentioned way of avoiding FI, namely, by appealing to the way in which the past, but not the future, is "over and done with." For instance, John Martin Fischer writes:

> To use an example from American football, the Atlanta Falcons "choked" terribly and lost the Super Bowl in the fourth quarter to the New England Patriots. I know that they would love to do something about this now; the Falcons would love to have access now to a possible world in which they did not lose the Super Bowl. But there is just nothing they can do about it, insofar as the game is now over-and-done-with. (2017, pp. 82-83)

According to this line of thought, every past fact is "over and done with," and any fact that is "over and done with" is beyond any agent's control. Hence, this proposal, if successful, would seem to explain the asymmetry between past and future without invoking anything like FI.

I am quite sympathetic to the suggestion that the past is "over and done with." But there are, again, at least two reasons why FI as an *explanation* of the asymmetry between past and future ought to be preferred to this suggestion. First, there is a powerful dilemma for this suggestion: either the notion of "over-and-done-with-ness" is understood in a temporally neutral way – i.e. in a way such that "over-and-done-with-ness" doesn't contain the notion of "pastness" – or it is not. If it is, then it is a coherent but unsatisfactory explanation for the asymmetry between past and future. If it isn't, then it is no explanation at all, but a mere restatement of the asymmetry between past and future. Let's take each horn of the dilemma in turn.

One popular way of understanding "over-and-done-with-ness" involves the notion of "temporal instrinsicality." For instance, John Martin Fischer and Neal Tognazzini claim it is plausible that "…fixity stems from over-and-done-with-ness, and over-and-done-with-ness is a function of temporal intrinsicality…" (2014, p.366).[18] What exactly is it for a fact to be temporally intrinsic? It is a difficult notion to pin down, but the basic idea is intuitive enough: a fact is temporally intrinsic if it only involves (or requires) events at a particular time. For instance, the fact that John F. Kennedy was assassinated in 1963 is commonly said to be a temporally intrinsic fact because it only involves the events of a particular time, namely the year 1963. In contrast, the fact that John F. Kennedy was assassinated 57 years prior to my writing this chapter is not a temporally intrinsic fact because it not only involves what happens in 1963 but also what happens in 2020, a time long after 1963.

If we understand "over-and-done-with-ness" solely in terms of temporal intrinsicality, then "over-and-done-with-ness" is a temporally neutral notion – the

[18] For clearly related but less lucid statements, see Pike (1965) and Hasker (1989).

notion of "over-and-done-with-ness" doesn't contain the notion of pastness. After all, it seems conceptually possible for there to be temporally intrinsic *future* facts. So far, so good. The problem is that now it is unclear what a fact's being "over-and-done-with" has to do with control. Simply because a (non-present) fact is temporally intrinsic, it does not follow that the fact is beyond our control. To see this, consider a temporally intrinsic future fact. Suppose it is a fact that Smith will be assassinated in 2063. If JFK's assassination counts as a temporally intrinsic fact, then surely Smith's assassination does as well. But plausibly, Smith's assassination is within our control: there seem to be steps that some of us could take to prevent Smith's assassination. (For example, Smith's parents could decide to not have any children.) Hence, temporal intrinsicality *by itself* doesn't have anything to do with control.

I suspect those sympathetic with the "over and done with" explanation will insist that it is not just a fact's being temporally intrinsic which accounts for its fixity, but rather its being temporally intrinsic *and past*. But this suggestion brings us to the second horn of the dilemma. Because this understanding of "over-and-done-with-ness" contains the notion of pastness, it seems that "over-and-done-with-ness" cannot genuinely *explain* the asymmetry between future and past. Under this understanding, to assert that any fact which is "over-and-done-with" is beyond our control *just is* to assert that the past, but not necessarily the future, is fixed. We can simply reframe the question we are interested in as follows: "Why does it seem that every temporally intrinsic past fact is fixed but not (necessarily) every temporally intrinsic future fact?" Plainly, the "over and done with" explanation does not answer that question. But this question does not seem substantively different than the question we started with, namely, "Why does it seem that the past is fixed but not (necessarily) the future?" Hence, the "over and done with" explanation does not answer the question we started with either.

So, it seems as if "over-and-done-with-ness" cannot explain the asymmetry between past and future. If we understand "over-and-done-with-ness" in a temporally neutral way, such as in terms of temporal intrinsicality, then "over-and-done-with-ness" is a coherent but unsatisfactory explanation. If we don't understand "over-and-done-with-ness" in a temporally neutral way, such as in terms of temporal intrinsicality *and* pastness, then "over-and-done-with-ness" is no explanation at all. Either way, FI is a better explanation for the asymmetry between past and future.

Even if I am wrong about this, though, there is still another reason to prefer FI as an explanation. Recall the second asymmetry mentioned above: that some of the future, but not all of it, is under our control. As we saw, FI has a straightforward explanation for this second asymmetry, namely, that some parts of the future, but not all of it, are within our causal reach. In contrast, it's hard to see how "over-and-done-with-ness" could explain this second asymmetry. Just as the growing-block theory says that *none* of the future exists, so it would seem that *none* of the future is "over and done with." How could "over-and-done-with-ness" explain why some of the future, but not all of it, is under our control? FI again seems to have more explanatory power.

To be clear, none of this is to say that the past *isn't* "over and done with." In fact, I am inclined to think of FI as being an explanation for, or even a partial precisification of, the past being "over and done with." But if "over-and-done-with-ness" is to be a *rival* explanation for the asymmetry between past and future, then it seems inferior to FI.

This leaves us with our third way of explaining the asymmetry between past and future while avoiding FI, namely, by adopting a principle substantially weaker than FI. In particular, consider the principle we started this section with: if a part of the world is beyond an agent's causal reach, then that part of the world is beyond the agent's control. I argued that this principle leads to FI, as it seems to be an instance of it. But perhaps this is mistaken. Why not simply accept this particular principle, but deny the more general principle of FI? That is, suppose someone accepted: (i) if a part of the world is beyond an agent's causal reach, then it is fixed for the agent; otherwise, it is not (necessarily) fixed for the agent. But suppose this same individual simultaneously denied: (ii) if a part of the world is beyond an agent's "explanatory" reach, then it is fixed for the agent; otherwise it is not (necessarily) fixed for the agent (i.e. FI). Such a position would easily explain why the past and some of the future is beyond our control. What pressure, then, is there to move beyond (i) and accept (ii)? Why move from causation to explanation? That is precisely the concern of the next section.

The Asymmetry Between the Hard and Soft Past

There is a (in)famous distinction between so-called "hard" facts about the past and so-called "soft" facts about the past. The following are often held to be hard facts about the past:

1. John F. Kennedy was assassinated in 1963.
2. The Atlanta Falcons lost Super Bowl 51.
3. Steven spent $500 last Saturday trying to repair his car.

Meanwhile, the following are often held to be soft facts about the past (supposing they are facts):

4. John F. Kennedy was assassinated 57 years prior to my writing this chapter.
5. The Atlanta Falcons lost their final Super Bowl appearance, Super Bowl 51. (Suppose the Falcons never make it to another Super Bowl.)
6. Steven wasted $500 last Saturday trying to repair his car.

Different authors have proposed different ways of distinguishing between hard facts and soft facts,[19] but we have already come across the basic idea: hard facts are "temporally intrinsic" facts about the past, whereas soft facts are "temporally extrinsic" facts about the past. As noted, (1) is plausibly a hard fact because it is solely about 1963; fact (2) is solely about Super Bowl 51; and fact (3) is solely about Steven's activities last Saturday. However, (4) is not solely about 1963 as it also

[19] The classic proposal is the "entailment criterion." See Adams (1967). More recently, Todd (2013) has given an importantly different proposal (and one that is much more plausible, to my mind at least).

involves my writing this chapter, something that happened long after 1963. Likewise, both facts (5) and (6) involve (or require) events that happen after the time in question: fact (5) requires that the Falcons never make it to the Super Bowl again, while fact (6) requires that Steven's car remain unrepaired.

Here is the interesting point: while hard facts about the past like (1) through (3) seem fixed for all of us once the relevant time has passed, soft facts about the past like (4) through (6) do not seem fixed for all of us once the relevant time has passed. Although (1) has always been fixed for me, (4) has not: presumably, I could have refrained from writing this chapter, and had I done so, (4) would not have obtained. And note that this seems correct *even if* (4) obtains in 1963. Likewise for (2) and (5): after the Falcons lost Super Bowl 51, there was nothing anyone could do to prevent (2) from obtaining, but there is plenty that could be done to prevent (5). (For example, the Falcons could win a future Super Bowl.) And similar comments apply for (3) and (6): once Steven spent the $500 on Saturday, there is nothing anyone can do to prevent (3), but presumably he or someone else could prevent (6). (If it was later discovered that a total of $600, say, was required to repair the car, Steven could pay an extra $100, and then his original $500 wouldn't have been wasted.)[20]

[20] Some, like van Inwagen (1983), have argued that facts (or propositions) are not true "at times" but rather "eternally" true. Hence, facts like (4) through (6) are not part of the past in any sense. But there are plenty of analogs to facts (4) through (6) that generate a similar issue. For instance, we could replace (4) with someone correctly believing in 1963 that I would write this chapter in 57 years. That part of the past doesn't seem fixed for the same reasons (4) doesn't.

Reflection on cases like these have convinced authors that FP ought to be restricted to the hard past. I agree. But now we have another asymmetry to explain: *why* does the hard past seemed beyond our control or fixed, but not (necessarily) the soft past? *Why* are facts like (1) through (3) fixed once the relevant time has passed but not, necessarily, facts like (4) through (6)?

FI again has a straightforward and powerful answer: at least generally, the hard past is in no way explained by any agent's current behavior whereas the soft past often is. Return to facts (1) through (6). Plausibly, the fact that Kennedy was assassinated in 1963 is in no way explained by my current behavior, certainly not my writing this chapter right now. But the fact that Kennedy was assassinated 57 years prior to my writing this chapter is explained, in part, by my writing this chapter right now. That is, fact (1) but not fact (4) seems explanatorily independent of my current behavior. Hence, FI implies that (1), but not necessarily (4), is fixed for me. Similar comments apply to the other facts. After the Falcons lost Super Bowl 51, no agent's behavior explains (2), but some agent's behavior does explain (5): the current and future behavior of the Falcon's organization seems to partly explain why the Falcon's lost their *final* Super Bowl appearance (supposing they never make it back to the Super Bowl, of course). Or after Steven spends his $500, no agent's behavior explains (3), but some agent's behavior does explain (6): Steven's failure to repair the car partly explains why Steven's money was not only spent, but *wasted*.

It's important to notice this, though: the way in which soft facts are explained by an agent’s later behavior does not seem to be a *causal* kind of explanation. By writing this chapter now, I am not *causing* fact (4) to obtain. It seems bizarre to think of an agent’s later behavior as causing some earlier soft fact to obtain, at least in the typical sense of “cause.” At the very least, such an admission would admit of rampant backwards causation. But if an agent’s later behavior does not cause earlier soft facts to obtain, then the weaker principle given in the previous section cannot explain why soft facts are not necessarily fixed. Recall the weaker principle: if a part of the world is not within the agent’s causal reach, then that part of the world is fixed for the agent; otherwise, it is not (necessarily) fixed for her. Since no agent’s current behavior *causes* soft facts to obtain, this principle cannot explain why the hard past but not the soft past is fixed. If anything, it seems to imply that the soft past *is* fixed for us precisely because the soft past is beyond our *causal* reach. Meanwhile, FI not only avoids this problematic implication, but also explains why the hard past, but not necessarily the soft past, is beyond our control.

So, reflection on hard and soft facts gives us strong reason to endorse FI, the more general principle, over a weaker principle. Once we notice that some soft facts are under our control, this suggests that instead of only focusing on what is *causally independent* of the agent’s behavior, we should broaden our scope and focus on what is *explanatorily independent* of the agent’s behavior as well.

In closing this section, I'd like to point out one more related phenomenon that FI can explain. It has been noted that, while *some* of the soft past seems to be under our control, not all of it is.[21] Consider once again fact (4): that Kennedy's assassination occurred 57 years prior to my writing this chapter. As noted, this soft fact does not seem fixed for me now (i.e. at the time of my writing this chapter). Compare this to fact (5): that the Atlanta Falcons lost their final Super Bowl appearance. Barring a drastic change in careers, this soft fact does seem fixed for me now – it doesn't seem at all *up to me* whether the Falcons ever make it back to the Super Bowl. There seems to be no action I can perform such that, had I performed it, (5) would (or might) not have obtained.

Given that both (4) and (5) are soft facts though, why should one be fixed for me and not the other? FI has a straightforward explanation: (4), but not (5), is partly explained by *my* current behavior (namely, my writing this chapter now). Fact (5), while being explained by some agent's behavior after Super Bowl 51, is in no way explained by *my* current behavior. Hence, FI implies that (5), but not necessarily (4), is fixed for me. FI not only explains why the hard past, but not necessarily the soft past, is fixed; it also explains why only *some* of the soft past appears to not be fixed.[22]

[21] See Fischer (1986) for discussion of this point.

[22] Or consider the fact that the sun's rising tomorrow is approximately 24 hours after its rising today. That seems like a soft fact beyond anyone's control. FI delivers this result because the sun's rising on any given day is not explained by anyone's behavior.

Before moving on to the next argument, let's briefly take stock. So far, I have argued that FI can explain two central asymmetries: the asymmetry between past and future, as well as the asymmetry between the hard and soft past. I've also argued that FI can explain some related phenomena: why only *some* of the future appears to be under our control, and why only *some* of the soft past appears to be under our control. It's important to note, though, that none of this calls FP into question. On the contrary, much of the foregoing *presupposes* that FP is at least generally correct. Instead, I have only purported to show that FI is the deeper and more general principle behind FP. That in itself is a significant result because, if correct, it would seem to suggest that when FP and FI deliver distinct results, FI is the principle to rely on. But it must be admitted that there is still an important option for those who accept FP. Such authors could "dig in their heels" and say that, regardless of whether FI is correct, FP is simply too intuitive to give up. It is precisely this option that is the focus of the next two arguments. The argument from relativity argues that FP is not all that intuitive, at least once it is reformulated in terms that are friendly to modern physics; the argument from time travel challenges the truth of FP.

FP, FI, and Relativity

Crucial to FP is the notion of "the past." But it is well known that "the past" is not a well-defined notion in standard interpretations of relativity.[23] Hence, if FP is to respect such interpretations, it must be reformulated in terms that are well-defined within the theory.

Now at first glance, this might seem like a small issue for FP as there are plenty of proxy notions of "the past" in relativity. Perhaps FP can be reformulated using one of these proxy notions. But the problem runs much deeper than this, for if FP is reformulated using one of these proxy notions, FP loses much of its intuitive appeal.

For instance, here is one way to reformulate FP in a relativistic setting: restrict FP to the facts (or events) *on or in the past light cone* of the agent's current behavior. Very roughly, a fact (or event) is on or in the past light cone of an agent's current behavior just in case something could travel from the fact (or event) to the agent's behavior without exceeding the speed of light. For example, Kennedy's assassination in 1963 is in the past light cone of our current behavior and, hence, fixed according to this restricted version of FP. Call FP so restricted "FP-CONE."

There are three problems with FP-CONE. First, it is far too weak. To see this, suppose that an astronomer, Galileo, starts scribbling in his notebook while

[23] See Geroch (1978) and Maudlin (2012) for accessible introductions to relativity theory. For more advanced treatments, I have found Einstein (1922/2014), Fleisch (2012), and Zee (2013) most helpful.

observing a distant star. He plans on observing the star for another hour. However, from Galileo's point of view, the star has already gone supernova thousands of years ago. If Galileo stays put and continues to observe the star, he will receive light from the supernova and witness its immediate destruction in thirty minutes. While the supernova is part of the distant past relative to Galileo's point of view, it lies just outside the past light cone of his current behavior. (This situation is illustrated in figure 1. The shaded area is the past light cone of Galileo's scribbling. The line labeled "Galileo's world line" represents Galileo's position in space at any given time.)

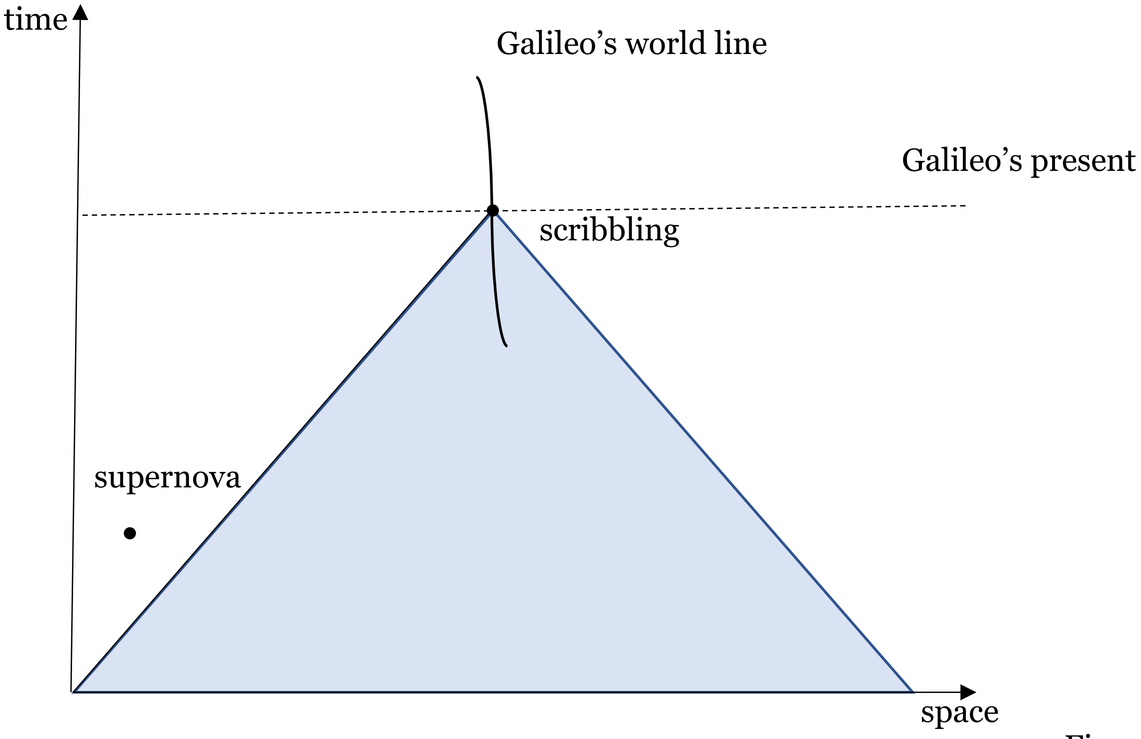

Fig. 1

Intuitively, the supernova is fixed for Galileo, being part of what Galileo would deem the distant past. And because of that, it would seem that Galileo cannot observe the star for another hour. But FP-CONE does not deliver these results. Given that the supernova is outside Galileo's past light cone, FP-CONE does not insist that the supernova is fixed and, hence, does not imply that Galileo cannot observe the star for another hour. It looks like FP-CONE is far too weak.

Upon reflection, it is easy to see why FP-CONE is unsatisfactory. An adequate formulation of FP should capture all of the facts that are "over and done with" for the agent (even if "over-and-done-with-ness" does not *explain* why such facts are fixed). Generally, facts that obtain thousands of years ago (from the agent's point of view) appear to be "over and done with" regardless of whether they obtain in the past light cone of the agent's current behavior. Therefore, it is unsurprising that FP-CONE is much too weak.

The second problem for FP-CONE is subtler. Given standard assumptions, it appears equivalent to a principle about the fixity of certain causal facts. It is commonly supposed that every cause must obtain (or occur) on or in the past light cone of its effect. If so, then FP-CONE appears equivalent to the claim that every fact that is *potentially causally relevant* to the agent's current behavior is fixed. This equivalence makes it dubious whether it is really the fixity of the past per se that undergirds FP-CONE, or whether it is the fixity of certain (potentially) causally relevant facts. If the latter, then FP-CONE is a short step away from FI. After all, assuming that causes are not explained by their effects, FI also implies that all causally relevant facts are fixed for the agent in question. Hence, this only lends further reason to think that FP, now understood as FP-CONE, is merely a derivative principle, and that FI is the deeper principle here.

The third problem for FP-CONE is that, absent some deeper explanation, it seems somewhat arbitrary. It is a fickle matter, in some sense, which facts lie on an agent's past light cone at a certain time. It is fickle because which set of facts lies on an agent's past light cone is determined by the agent's spatial position (at that time). To return to Galileo's case, while the supernova is not on his past light cone, it could be on his friend's past light cone who is watching the supernova from down the road (to use an unrealistic example). FP-CONE implies that the supernova is fixed for Galileo's friend, but not (necessarily) for Galileo. And this is so *despite* the fact that both would deem the supernova as having occurred at the same time, thousands of years ago in the past.

Cases like this make the explanatory question from the beginning of this chapter all the more pressing. Recall that we first asked: why is the past fixed but not, necessarily, the future? Under FP-CONE, this question becomes: why are the facts on my past light cone fixed but not, necessarily, the facts outside my past light cone? And that question seems to demand an answer. It is not clear what a fact's being inside or outside my past light cone *itself* has to do with its being (possibly) under my control. And insofar as this revised question seems more pressing, that seems to suggest that FP-CONE lacks intuitive appeal.

So, FP-CONE is far from satisfactory. Let's try another way of modifying FP: restrict FP to all of the facts deemed past *from the agent's frame of reference*. Very roughly, an agent's frame of reference is a coordinate system (of a certain type) that the observer can use to measure spatial and temporal distances between facts (or events). Call this suggestion ”FP-FRAME.” This suggestion seems to avoid the first two problems that FP-CONE faces. First, it gets the right result both in regard to Kennedy's assassination and Galileo's supernova. Kennedy's assassination is deemed past from our current frame of reference and, hence, fixed for us; the supernova is deemed past from Galileo's frame and, hence, fixed for him. (To refer to figure 1, FP-FRAME implies that any fact beneath the dashed line is fixed for Galileo.) More generally, FP-FRAME seems to do a better job of capturing the facts that are "over and done with" for the agent. And it also seems to avoid the second problem because it is not equivalent to any principle about the fixity of certain causal facts.

However, FP-FRAME does face its own version of the third problem. Just like FP-CONE, FP-FRAME also appears somewhat arbitrary. Which set of facts an agent deems as past is also a fickle matter because it is determined by the agent's velocity. While Galileo might deem the supernova as past, a friend who passes him by very quickly on the street might deem it as future (to use another unrealistic example). FP-FRAME implies that the supernova is fixed for Galileo but not (necessarily) his friend. That might seem puzzling initially.

Let's make it more puzzling, though. Suppose that before he starts scribbling in his notebook, Galileo decides to take a relatively short trip in a spaceship. With this change in velocity, Galileo comes to occupy a different (inertial) frame of reference, one where he no longer deems the supernova as past. In this case, FP-FRAME implies that the supernova is fixed for Galileo before he boards the spaceship, but that it is not (necessarily) fixed for him after his journey has begun. Furthermore, once Galileo's short trip is over and he comes to occupy his original (inertial) frame of reference again on Earth, FP-FRAME implies that the supernova is once again fixed for Galileo. (The first leg of Galileo's trip is represented in figure 2, where the sharp turn in his world line represents his spaceship taking off.)

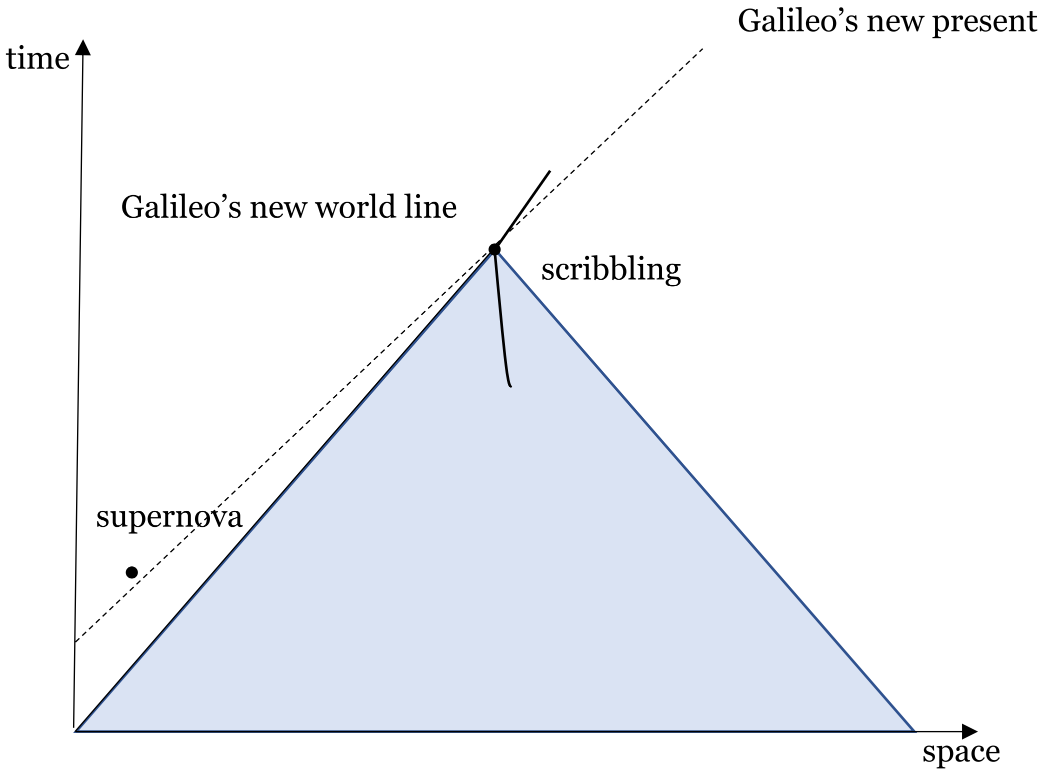

Fig. 2

All of this seems bizarre. FP-FRAME implies that the supernova goes from being fixed, to not necessarily fixed, and back to fixed again all because Galileo took a short trip in a spaceship. Just like with FP-CONE, cases like these make the explanatory question all the more pressing. Under FP-FRAME, the question is: why are the facts deemed past from my frame of reference fixed but not, necessarily, the facts deemed future from my frame of reference? Once we see that an agent's frame of reference is determined by the agent's velocity, this question demands an answer. It is not clear what a fact's being past from my frame of reference *itself* has to do with its being (possibly) under my control. FP-FRAME simply lacks intuitive appeal.

There is a familiar point worth mentioning here. Again, nothing I have said here implies that FP-CONE or FP-FRAME is *false*. On the contrary, it is conceptually possible that FI implies both FP-CONE and FP-FRAME. Instead, I take the challenge of relativity to be a challenge to the intuitive appeal of FP. In a Newtonian setting, FP does seem quite intuitive or even obvious. When asked why the past, but not necessarily the future, is beyond our control, many give blank stares. "It's beyond our control because it's the past, obviously!" That is, the intuition behind FP seems so powerful to some that trying to explain the asymmetry between past and future might seem like an insignificant endeavor.

Once we shift to a relativistic setting though, FP (or some restriction thereof) does not seem so intuitive. It is not obvious why (or even *that*) the facts on or in my past light cone, say, are fixed for me but not (necessarily) those outside my past light cone. FP-CONE and FP-FRAME need some motivation. And insofar as they do, that seems to show that they both lack intuitive appeal. At the very least, they are not nearly as intuitive as FP in a Newtonian setting.[24]

Contrast this with FI and relativity. First, FI requires no reformulation because it does not make use of any notion, like "past" or "prior," that is not well-defined in relativity. Second, FI delivers intuitive results. To return to Galileo's cases, plausibly the supernova is in no way explained by Galileo's nor his friends' current behavior. FI therefore implies that the supernova is fixed for Galileo and his friends *regardless* of their spatial positions and frames of reference. More generally, the shift to relativity does not seem to undermine the intuitive appeal of FI or its implications whatsoever. That is a significant point in favor of FI.[25]

So the shift from a Newtonian conception of time to a relativistic one only gives us more reason to think of FP (or some restriction thereof) as derivative of FI. But in closing this section, let's consider one last suggestion that those

[24] Some might suggest that relativistic variants of FP lack intuitive appeal simply because the theory of relativity is so unintuitive. Blame the physics, not the metaphysics! But this seems mistaken, at least in my own case. Even after I have let the strange physics "settle in," relativistic variants of FP still seem far from obvious to me.

[25] What if FP was restricted to the facts which obtain outside the *future* light cone of the agent's current behavior? First, it is somewhat strange to call this a version of the fixity of the *past*, as many of the facts that obtain outside the future light cone of an agent's current behavior would be deemed (by some) as occurring long after the agent's behavior. It would be better to call it the principle of the "fixity of the not-absolute-future." But second, and more importantly, this restriction plays right into the hands of FI. It is commonly supposed that an agent's behavior can only causally affect those facts on or in the future light cone of the agent's behavior. If this supposition is correct, then this restriction is simply an instance of FI. And if this supposition isn't correct, then it isn't clear why we should accept this restriction – what does a fact's being outside of my future light cone *itself* have to do with whether it is fixed for me? Either way, this restriction seems inferior to FI. Thank you to Neal Tognazzini for bringing up this restriction.

sympathetic to FP might give: reject standard interpretations of relativity. For example, many *A*-theorists embrace non-standard interpretations of relativity that allow for absolute notions of the past, present, and future. Perhaps FP could follow suit and not require any reformulation at all.

These are deep waters, but there are two reasons why FI is to be preferred here. First, and most obviously, non-standard interpretations of relativity are extremely controversial. Insofar as FI does not need to take a stand on this issue, but FP does, that is a point for FI. Second, even setting aside the controversy, it is still not clear that non-standard interpretations can save the intuitive appeal of FP. In short, non-standard interpretations of relativity typically privilege some point in spacetime, frame of reference, or foliation of hyperplanes.[26] The past, present, and future are then defined relative to the privileged entity. However, all parties admit that it is extremely unlikely that the privileged entity agrees with us (exactly) on what is past, present, and future. There will almost certainly be facts we would deem past and "over and done with" that are actually future and vice versa. To use Galileo's case, it is possible that Galileo deem the supernova as part of the distant past, but it actually be part of the future. (For example, it could be that the dashed line in figure 1 represents what Galileo deems as present, while the dashed line in figure 2 represents the *real* present.) But surely, the supernova is "over and done with," or at least fixed, for Galileo. Thus, under this suggestion, FP fails to capture what FP ought to capture, namely, all of the facts that we would deem as "over and

[26] See Godfrey-Smith (1979), Forrest (2008), and Crisp (2007), respectively.

done with." So, even if these *A*-theorists are correct to reject standard interpretations of relativity (which is highly contentious), doing so will not help save the intuitive appeal of FP.

So however we revise FP to accommodate modern physics, it simply lacks much intuitive support. This makes the response considered at the end of the last section implausible. It is one thing to "dig in one's heels" for a highly intuitive principle; it is another to do so for one that is not. And now we are in a position to appreciate the final argument against FP and for FI.

FP, FI, and Time Travel

The final argument for FI appeals to time travel. Now, I am not the first to argue for FI over FP by appealing to time travel. For instance, one of the authors previously mentioned, Philip Swenson (2016), has given such an argument. He writes:

> Imagine that you have come to believe that you are sitting in a working time machine... You believe that the machine is programmed so that, if you push the button in front of you, then you will travel to the year 1492. Furthermore, you believe that the past and the laws entail that you will travel to 1492 if and only if you push the button. (p. 664)

Suppose you push the button and travel to 1492. Swenson claims that, intuitively, at the moment of your decision, you can both push and refrain from pushing the button. But it's hard to see how FP could accommodate this intuition given that your appearance in 1492 is past and, hence, fixed according to FP. By comparison, FI can accommodate this intuition because your appearance in 1492 is partly explained by your pushing the button and, hence, not (necessarily) fixed. That seems like a point in favor of FI.

I find Swenson's argument unconvincing. First, in my experience, those who prefer FP over FI often don't have the intuition that you can both push and refrain from pushing the button. And even for those who do, there seems to be an adequate error theory here. The way Swenson describes the case, you do not know whether you appear in 1492 at the time of your decision and, thus, it is a genuine epistemic possibility that you refrain from pushing the button. Given that we often conflate epistemic possibilities with metaphysical possibilities, you might be tempted to think that you can refrain. But epistemic possibilities are not always metaphysical possibilities. In fact, if we redescribe the case in such a way that it is *not* an epistemic possibility for you to refrain from pushing the button, it is no longer clear that you are free to refrain. For instance, suppose that before the moment of your decision, you come across an old journal from 1492 detailing the arrival of a mysterious figure who claimed to be from the future. This person is described as having your name and exact appearance. Moreover, this person made many correct and striking predictions about the future, some particular to your life – the names of your parents, your birth place, your best friend's name from the third grade, etc. You become convinced that this mysterious figure was you.

Once you become convinced of this, is it still so obvious that you are free to refrain from pushing the button? The intuition seems to evaporate, or so the advocate of FP can claim. If so, then it would seem that Swenson's case is only compelling because we mistakenly conflate an epistemic possibility with a metaphysical one. Once the epistemic possibility is taken away, it is clear that the metaphysical possibility was never there to begin with.

By my lights, this error theory is an adequate response to Swenson's argument. Hence, this case of time travel doesn't seem to provide a compelling argument for FI. At best, it serves as a good test case for whether one ought to endorse FP or FI. If one *still* has the intuition that you are free to push the button, even when the epistemic possibility has been taken away, then one ought to endorse FI instead of FP. But this doesn't seem to be much of an *argument*.

My argument from time travel is not so easily dismissed. Here is the argument in a nutshell. We can distinguish between two types of (backward) time travel cases. First, there are cases where the time traveler is caught in an explanatory loop. Second, there are cases where the time traveler isn't caught in an explanatory loop. I claim that our intuitions about what the time traveler can and can't do varies across these different cases. And while FI, or something like it, can explain this difference in intuition, FP cannot. Moreover, there is no obvious error theory that can explain away the difference in our intuitions.

Let's start with a case of a time traveler caught in an explanatory loop. Suppose Bill stumbles upon some plans for a time machine in his family's attic and uses them to build a time machine in 2020. He then travels back to 1900, gives his great-great-grandfather the plans, but dies of influenza shortly thereafter. His great-great-grandfather puts the plans in the attic, where he eventually forgets about them. Years later, in 2020, (young) Bill stumbles upon the plans, uses them to build his time machine, travels back to 1900, and so on and so forth.

For many, especially those sympathetic to FP, it is quite intuitive that Bill cannot refrain from pushing the button in this case.[27] After all, if he were to refrain, the history of the world would be significantly different, maybe even contradictory. (Where would the plans have come from? How could he have built a time machine?) And while FP seems to get the right result here, it looks as if FI doesn't. Plausibly, the fact that Bill appeared in 1900 is dependent on his pushing the button. According to FI, that means his appearance in 1900 is not (necessarily) fixed for him in 2020 and, thus, there is no reason to think he cannot refrain from pushing the button.

Fortunately, there is a fairly straightforward way of amending FI to accommodate this intuition. FI claims that any part of the world which is not explained by the agent's behavior is fixed for the agent. But a natural principle to pair with FI is that any part of the world which *explains the agent's behavior* is

[27] A fair amount of ink has been spilled about freedom and time travel, and I cannot engage with all of it here. Lewis (1976) contains the classic discussion; see Wasserman (2018, chs. 3 and 4) for a recent overview.

also fixed for the agent.[28] If fact *F* partly explains why an agent performed a certain action, it might seem illicit to not hold *F* fixed in determining whether she could have done otherwise. For example, if part of the explanation for why you remained seated was because you were strapped to the chair, it would seem illicit to not hold that fact fixed in determining whether you could have stood up. If this additional principle is accepted, we can augment FI as follows:

> FI+: Agent *S* can perform action *X* at time *t* (in world *w*) only if there is a world, *w'*, such that (i) all of the facts that are distinct from and explanatorily independent of the facts constituting *S*'s behavior at time *t* hold, (ii) all of the facts that *S*'s behavior at time *t* explanatorily depend on also hold, and (iii) *S* performs *X* at *t*.[29]

FI+ seems to imply that time travelers caught in explanatory loops, like Bill, are not free. Given that his appearance in 1900 partly explains (at least ancestrally) his pushing the button on his time machine in 2020, clause (ii) insists that his appearance in 1900 is fixed for him in 2020. Assuming there is no world where he refrains from pushing the button but still appears in 1900, FI+ implies that Bill is not free to refrain from pushing the button.[30]

[28] For instance, see Rea (2015). Rea's paper is especially fitting given that he is explicitly concerned with time travel.

[29] One might worry that FI+ is too strong as it straightforwardly implies that determinism and the freedom to do otherwise are incompatible. See the end of chapter 3 for discussion.

[30] For ease of exposition, I am assuming that explanation is a transitive notion. But even if it is not, clause (ii) of FI+ could easily be amended in such a way that any fact which stands in the "ancestral explanatory" relation is fixed. That would deliver the same result.

It is important to note that FI+ isn't too much of a leap from FI. The only difference between FI and FI+ is that the latter adds clause (ii). But in the vast majority of cases, and perhaps all actual ones, the facts in clause (i) will encompass the facts in clause (ii). That is, in the vast majority of cases, if fact *F* explains an agent's current behavior, then the agent's current behavior does not explain *F*. Hence, holding fixed all of the facts that are not explained by the agent's current behavior will almost always require holding fixed all of the facts that explain the agent's current behavior. Time travel cases with explanatory loops are the exception, not the rule.

So, it may be fairly intuitive that time travelers caught in explanatory loops are not free, at least for those drawn to FP in the first place. Fortunately, FI (or FI+) can accommodate this intuition. But now consider a case of a time traveler not caught in an explanatory loop.[31] Suppose Ted stumbles upon some plans for a time machine in his family's attic. However, these plans came about in a more ordinary way: his great-great-grandfather developed the plans on his own, put them in his attic, but eventually forgot about them. Ted uses the plans to build his time machine in 2020. He then travels not just to a distant time, but a distant place as well: he travels to the year 1900 but on a barren planet, galaxies away. Upon arrival, Ted and his time machine are immediately annihilated, leaving no trace of his (not-so-excellent) adventure.

[31] There is some controversy over whether it is possible for backward time travel to occur without creating a causal (and, hence, explanatory) loop. See Monton (2009) for a defense of the claim that it is possible.

Can Ted refrain from pushing the button on his time machine? I see no reason to think not. If he hadn't pushed the button, he still would have had access to his great-great-grandfather's plans, he still would have had the same personal past up until that moment, the history of the world (until 2020) would not have been significantly different, and no obvious contradiction would rear its head. At the very least, it does not seem as intuitive that Ted *must* push the button in this case.

FI (and FI+) is compatible with the intuition that Ted is free to refrain from pushing the button. His pushing the button plainly explains his brief appearance in the past. So, clause (i) of FI+ does not insist that his appearance is fixed. And in contrast to Bill's case, his appearance in 1900 does not explain (even ancestrally) his pushing the button in 2020. So, clause (ii) does not insist that his appearance is fixed either. That means FI+ does not imply that his appearance in 1900 is fixed and, hence, is compatible with his being free to refrain from pressing the button in 2020.

More generally, there seems to be yet another asymmetry here. In cases where the time traveler is caught in an explanatory loop, it may be fairly intuitive that the time traveler isn't free; in cases where the time traveler is not caught in an explanatory loop, it isn't so intuitive that the time traveler isn't free. What explains this asymmetry?

It's hard to see how FP could explain this asymmetry, or if it is even compatible with it. In both kinds of cases, the time traveler's appearance is part of the past. What principled reason, within the spirit of FP, could there be for treating the two cases differently? Moreover, there is no obvious error theory for our intuitions here. For instance, the error theory I gave for Swenson's case, about conflating an epistemic possibility with a metaphysical possibility, certainly won't do. First of all, the case is not first-personal. But more importantly, Bill's and Ted's epistemic possibilities seem to be exactly the same in both cases. Absent some forthcoming error theory, FP seems deficient here.

However, FI (or FI+) seems perfectly suited to explain the difference between these two kinds of cases. Given that the difference between the two types of cases is precisely a difference in explanatory structure, it is unsurprising that FI (or something similar) could capture and explain the asymmetry. FI has the advantage here.

Notice that this is the first case I have given which implies that FP delivers the *wrong* results. If Ted really is free to refrain from the pushing the button on his time machine, then *contra* FP, his appearance in the past is not fixed at the time of his decision. But I want to emphasize that these kinds of cases are far from the norm (or even actuality). Hence, even if cases like these show that FP is false, it still seems to be *approximately* (or perhaps even contingently) correct.

Concluding Remarks

This concludes my case for why FI ought to be preferred to FP. In short, FI can explain not only the intuitive appeal of FP, but much more. Moreover, it does so in an ecumenical way, without requiring terribly controversial commitments. All else being equal, this gives us good reason to think that FI is the deeper principle behind FP, and that FP is at best a derivative and contingent principle. Hence, if FI and FP were to deliver distinct results over a particular issue, like they seem to in the case of divine foreknowledge, the deliverances of FI ought to be preferred to the deliverances of FP.

In closing, I'd like to offer an analogy to help illustrate the position articulated in this chapter. Astrophysicists have been looking for extraterrestrial life for some time and they have recently discovered a planet which seems promising. The planet is in the so-called "goldilocks" zone, being just the right distance from its sun to have temperatures that would allow for liquid water on its surface. But there is a problem. Just like our moon, the planet is "tidally locked," meaning that it does not rotate on its axis. One of its hemispheres is continually facing its sun, presumably resulting in extremely hot temperatures, while the other hemisphere is continually facing away from its sun, presumably resulting in extremely cold temperatures. If there is going to be life found on this planet, it is probably going to be found where one hemisphere meets the other, right in-between the extreme warmth and the extreme cold.[32]

Suppose we find intelligent life like our own on this planet. You become so excited upon learning this that you decide to visit, not knowing much of anything about the planet other than that it houses intelligent life. Upon arriving, you learn that the inhabitants have very different attitudes about the two hemispheres, despite the fact that they only currently live in-between the two. While they feel very optimistic about the eastern hemisphere, the cold one, they feel that the western hemisphere, the hot one, is a lost cause. They say things like: “Don’t dwell on the western hemisphere. All you can do is plan for the eastern,” and “It’s no use crying over milk left in the western hemisphere.”

A particular interaction drives home how deep this asymmetry in attitudes is. One day during your stay, the inhabitants learn that something terrible is going to happen somewhere on the planet. At first, there are concerted efforts to try and prevent this terrible event, but eventually the inhabitants also come to learn that this event will take place in the western hemisphere, and so they resign themselves to the event’s occurrence. Naturally, you ask them why they have given up. In response, they simply say: “It’s in the western hemisphere. If it had been in the eastern hemisphere, perhaps we could have done something. But what’s in the west is in the west.”

[32] See Anglada-Escudé, *et al.* (2016).

At first, you are baffled. Why do the inhabitants have this asymmetry in attitudes about the western and eastern hemispheres? Upon further research, you find your answer. The western hemisphere is so hostile that it is utterly impossible to affect what goes on there. Many have tried, but no one has even set a foot in the western hemisphere without immediately dying of dehydration, radiation poisoning, and all the rest that the intense rays of the sun bring. In contrast, some inhabitants have been able to affect the eastern hemisphere. While it's difficult to have any kind of lasting effect, with some hard work and perseverance, changing the eastern hemisphere seems possible.

Here is the interesting point. When you ask the inhabitants why they have resigned themselves to the occurrence of the terrible event, it is in no way informative *for you* to be told that the event's occurrence lies in a certain spatial direction – it is not informative *for you* to be told that "What's in the west is in the west." For *the inhabitants* who have lived their entire lives on this planet, being told as much might very well be informative. But that is only because they implicitly know that what lies in the west is beyond their causal reach and, hence, beyond their control. It is clear that the spatial direction *itself* has nothing to do with the event's being beyond their control. And if they were to discover, perhaps *per impossible*, that a certain portion of the western hemisphere is within their causal reach, they should reconsider whether the portion in question is beyond their control.

It seems to me that we are in an analogous position with regard to *time*. When we learn that an event's occurrence lies in a certain *temporal* direction, namely the past, we resign ourselves to the event's occurrence – "What's in the past is in the past." But just as spatial direction *itself* has nothing to do with an event's being beyond our control, so I have suggested that temporal direction *itself* has nothing to do with an event's being beyond our control. If learning that a certain event is past seems relevant to determining whether the event is under our control, that is only because we implicitly know that what lies in the past is beyond our causal (or explanatory) reach and, hence, beyond our control. And if we were to discover, perhaps *per impossible*, that a certain portion of the past is within our causal (or explanatory) reach, we should reconsider whether the portion in question is beyond our control. Or so I have argued.

References:

Adams, Marilyn M. (1967). “Is the Existence of God a ‘Hard Fact’?” *Philosophical Review* 76: 492-503.

Anglada-Escudé, G. *et al.* (2016). “A Terrestrial Candidate Planet in a Temperate Orbit Around Proxima Centauri,” *Nature* 536: 437-440.

Aristotle (350 BCE/ 1963). *De Interpretatione*, translated by Ackrill, J.H. Oxford: Clarendon Press.

Briggs, Rachael & Forbes, Graeme A. (2012). “The Real Truth about the Unreal Future,” in *Oxford Studies in Metaphysics*, vol. 7., edited by Karen Bennett and Dean Zimmerman. Oxford: Oxford University Press: 257-304.

Crisp, Thomas M. (2007). “Presentism, Eternalism, and Relativity Physics,” in *Einstein, Relativity, and Absolute Simultaneity*, edited by William L. Craig and Quentin Smith. New York: Routledge: 262-278.

Dorr, Cian (2016). “Against Counterfactual Miracles,” *Philosophical Review* 125: 241-286.

Einstein, Albert (1922/2014). *The Meaning of Relativity*. Princeton: Princeton University Press.

Finch, Alicia & Rea, Michael (2008). “Presentism and Ockham’s Way Out,” in *Oxford Studies in Philosophy of Religion*, vol. 1, edited by Jon L. Kvanvig. Oxford: Oxford University Press: 1-17.

Fischer, John M. (1986). “Hard-Type Soft Facts,” *Philosophical Review* 95: 591-601.

Fischer, John M. (1994). *The Metaphysics of Free Will: An Essay on Control*. Malden: Blackwell Publishing.

Fischer, John M. (2011). “Foreknowledge, Freedom, and the Fixity of the Past,” *Philosophia* 39: 461-474.

Fischer, John M. (2017). “Replies to my Critics,” *European Journal for Philosophy of Religion* 9: 63-85.

Fischer, John M. & Todd, Patrick (2011). “The Truth about Freedom: A Reply to Merricks,” *Philosophical Review* 120: 97-115.

Fischer, John M. & Tognazzini, Neal A. (2014). “Omniscience, Freedom, and Dependence,” *Philosophy and Phenomenological Research* 88: 346-367.

Fleisch, Daniel (2012). *A Student’s Guide to Vectors and Tensors*. New York: Cambridge University Press.

Forrest, Peter (2006). “Uniform Grounding of Truth and the Growing Block Theory: A Reply to Heathwood,” *Analysis* 66: 161-163.

Forrest, Peter (2008). “Relativity, the Passage of Time, and the Cosmic Clock,” in *The Ontology of Spacetime*, edited by Dennis Dieks. Oxford: Elsevier: 245-254.

Geroch, Robert (1978). *General Relativity: From A to B*. Chicago: University of Chicago Press.

Godfrey-Smith, William (1979). “Special Relativity and the Present,” *Philosophical Studies* 36: 233-244.

Hasker, William (1989). *God, Time, and Knowledge*. Ithaca: Cornell University Press.

Holliday, Wesley H. (2012). "Freedom and the Fixity of the Past," *Philosophical Review* 121: 179-207.

Lewis, David (1976). "The Paradoxes of Time Travel," *American Philosophical Quarterly* 13: 145-152.

Maudlin, Tim (2012). *Philosophy of Physics: Space and Time*. Princeton: Princeton University Press.

McCall, Storrs (2011). "The Supervenience of Truth: Free Will and Omniscience," *Analysis* 71: 501-506.

Mellor, David H. (1998). *Real Time II*. London: Routledge.

Merricks, Trenton (2006). "Good-bye Growing Block," in *Oxford Studies in Metaphysics*, vol. 2, edited by Dean Zimmerman. Oxford: Oxford University Press: 103-110.

Merricks, Trenton (2009). "Truth and Freedom," *Philosophical Review* 118: 29-57.

Merricks, Trenton (2011). "Foreknowledge and Freedom," *Philosophical Review* 120: 567-586.

Monton, Bradley (2009). "Time Travel Without Causal Loops," *The Philosophical Quarterly* 59: 54-67.

Pike, Nelson (1965). "Divine Omniscience and Voluntary Action," *Philosophical Review* 74: 27-46.

Rea, Michael (2015). "Time Travelers are not Free," *Journal of Philosophy* 112: 266-279.

Sartorio, Carolina (2015). "The Problem of Determinism and Free Will Is Not the Problem of Determinism and Free Will," in *Surrounding Free Will: Philosophy, Psychology, and Neuroscience*, edited by Alfred Mele. Oxford: Oxford University Press: 255-273.

Sider, Theodore (2001). *Four-Dimensionalism: an Ontology of Persistence and Time*. Oxford: Oxford University Press.

Swenson, Philip (2016). "Ability, Foreknowledge, and Explanatory Dependence," *Australasian Journal of Philosophy* 94: 658-671.

Todd, Patrick (2013). "Soft Facts and Ontological Dependence," *Philosophical Studies* 164: 829-844.

Todd, Patrick & Fischer, John M. (2013). "The Truth about Foreknowledge," *Faith and Philosophy* 30: 286-301.

Tooley, Michael (1997). *Time, Tense, and Causation*. Oxford: Oxford University Press.

van Inwagen, Peter (1983). *An Essay on Free Will*. Oxford: Oxford University Press.

Wasserman, Ryan. (2018). *Paradoxes of Time Travel*. Oxford: Oxford University Press.

Westphal, Jonathan (2011). "The Compatibility of Divine Foreknowledge and Free Will," *Analysis* 71: 246-252.

Zee, Anthony (2013). *Einstein Gravity in a Nutshell*. Princeton: Princeton University Press.

Chapter 2: Objections to the Fixity of the Independent

Introduction

In the previous chapter, I provided my positive case for FI. Roughly, the argument went like this: it is quite intuitive that the past is beyond our control or "fixed" for us. The best explanation for this is that the past is beyond our explanatory reach, and that any part of the world beyond our explanatory reach is beyond our control or "fixed" for us. But this latter claim, that any part of the world beyond our explanatory reach is beyond our control or "fixed" for us, just is the principle of FI.

Like any inference to the best explanation though, this argument is only compelling insofar as the explanation on offer, in this case FI, doesn't face any serious problems of its own. So even if the arguments of the last chapter are compelling, it might be that, on balance, FI ought to be rejected. The goal of this chapter is to show that there are no serious problems with FI. While there are many objections to FI, I will consider and respond to those I find most pressing.

The Symmetric Universe Objection

Consider the following case, inspired by David Chalmers:

> **Symmetric Universe**: Dave lives in a universe very similar to ours but with one important difference: he lives in a perfectly symmetric universe. Dave's universe is split into two, with each half of the universe being a perfect "mirror image" of the other. Moreover, it is a law of physics that the universe remain perfectly symmetric. Hence, not only does Dave exist in this universe, but so does David, someone in the other half of the universe who is qualitatively identical to Dave. One day, Dave is deliberating about whether or not to raise his hand to ask his teacher a question. After weighing the reasons, he decides to do so and raises his hand at time *t*. Given the universe Dave lives in, David goes through the exact same process and ends up raising his hand at time *t* to ask his teacher the same question as well.

Intuitively, Dave's action of raising his hand at time *t* is a free action. But FI seems to disagree. Presumably, Dave's raising his hand at time *t* does not explain David's raising his hand at time *t*. To say otherwise would be problematic. If Dave's action does explain David's, then it would seem that the explanation must go the other way as well: that David's action also explains Dave's. After all, there doesn't seem to be any reason for privileging Dave's action over David's. But if Dave's action explains David's and *vice versa*, then the asymmetry of explanation is violated. That's clearly a problem. So, it seems as if we must admit that Dave's action at time *t* does not explain David's nor *vice versa*: we should admit that Dave's and David's actions are *explanatorily independent* of each other.

If Dave's action at time *t* in no way explains David's action at time *t*, then FI insists that David's action is fixed for Dave at time *t*. If we hold fixed the laws as well (which seem explanatorily independent of Dave's action at time *t*), then it would seem as if there is no world where Dave refrains from raising his hand at time *t*: every world where Dave so refrains is a world where David also refrains or where the laws of physics do not require the universe to be perfectly symmetric. Thus, FI implies that Dave is not free to refrain from raising his hand. That's the wrong result.[1]

The central difficulty with this objection, in my view, is that it conflates *freely* performing an action and *being free to do otherwise* than that action. It may indeed be that Dave freely raises his hand, but it is less clear that he is free to do otherwise than raise his hand. And insofar as FI only implies that Dave is not free to do otherwise, this case is not a significant challenge to FI. Allow me to elaborate.

Let's start by considering another case:

> **Locked Room**: Smith is in a large room and deliberating about whether or not to leave. She knows she has other things to eventually get to, but the room does have some pretty enticing stuff in it. There are many exquisite paintings, books, and other items of great interest to her. After thinking it over a while, she decides to stay in the room for a bit longer. However, unbeknownst to Smith, the room has been

[1] Thank you to David Chalmers for raising this objection during the 2018 USC/UCLA Graduate Student Conference.

> locked from the outside the entire time. Even had she wanted to leave, she would have remained in the room.[2]

Did Smith freely stay in the room? On the one hand, it's not as if she was forced to stay in the room against her will. She deliberated, weighed the reasons, and decided on her own to stay in the room. There seems to be some legitimate sense in which she freely stayed in the room. But on the other hand, the door was locked! She couldn't have left the room no matter how hard she tried. So there also seems to be some sense in which she did not freely stay in the room.

To capture these two different senses, I will say that Smith freely stayed in the room but that she was not free to do otherwise. Roughly, an agent freely performs action *X* if the agent intentionally performs *X*, isn't coerced or manipulated into performing *X*, performs *X* for reasons, etc. Cases like **Locked Room** seem to show that an agent can freely perform action *X* even if the agent isn't free to do otherwise than *X*: it is possible for an agent to intentionally perform *X*, not be coerced or manipulated into doing so, etc., and yet not be free to do otherwise than *X*.[3]

With this distinction, we can now return to the case of **Symmetric Universe**. Clearly, Dave freely raises his hand – he intentionally raises his hand,

[2] This case is a slightly modified version of John Locke's famous case, found in his *An Essay Concerning Human Understanding*, Book II, chapter XXI, section 10.

[3] So-called "Frankfurt-style" cases also putatively show that an agent can freely perform an action without being free to do otherwise. Indeed, Frankfurt-style cases supposedly show something stronger, namely, that the freedom to do otherwise is not required for moral responsibility. However, Frankfurt-style cases are controversial and, hence, I will be avoiding them. Moreover, I will remain neutral on whether the freedom to do otherwise is required for moral responsibility.

is not coerced or manipulated into doing so, does so for reasons, etc. But is it so clear that Dave is free to do otherwise than raise his hand? Here is one reason to think not: in order for Dave to do otherwise, David would also have to do otherwise. But David is billions of light years away, far beyond Dave's control. So, it seems as if Dave couldn't have done otherwise.

Whether or not this reason is conclusive, it does seem at least somewhat compelling.[4] Hence, it at least seems unclear whether or not Dave is free to do otherwise than raise his hand. And insofar as it is unclear, FI's implying that Dave is not free to do otherwise is not an obvious problem for FI.

In closing this section, I would like to point out an interesting difference between **Symmetric Universe** and **Locked Room**, one that might be thought relevant to Dave's and Smith's freedom. In **Locked Room**, if Smith had tried to leave the room, she would have failed – the lock would have stopped her. However, in **Symmetric Universe**, if Dave had tried to keep his hand still, he would have succeeded – presumably, nothing would have stopped him. (Of course, if Dave had kept his hand still, David would have also.) Some might think this difference shows that Dave, but not Smith, is free to do otherwise.

This line of reasoning clearly presupposes something like the following, often called the "Simple Conditional Analysis":

[4] This kind of reasoning is sometimes called "consequence-style reasoning." See van Inwagen (1983). And while many reject it, most admit that it has some intuitive appeal. Insofar as it does, I think it's fair to say that it is not *obvious* that Dave is free to do otherwise.

> **Simple Conditional Analysis**: Agent *S* is free to do otherwise than action *X* if, and only if, if *S* were to try to do otherwise than *X*, *S* would do otherwise than *X*.[5]

While the Simple Conditional Analysis was quite popular at one time, it is now widely rejected. The relevant problem with this analysis is that it does not provide a sufficient condition on the freedom to do otherwise. There seem to be cases where *S* is not free to do otherwise than *X* despite the fact that *if S* had tried to do otherwise, *S* would have succeeded. Here is one such case:

> **Spider**: Erik is at the zoo when an employee pulls out a tarantula, offering any brave visitor the chance to hold it. Erik, having a severe case of arachnophobia, runs away and bursts into tears.[6]

Intuitively, it is not within Erik's power to hold the spider. Given his crippling fear of spiders, it seems as if his own psychology denies him the freedom to do so. But now consider the following counterfactual: *if* Erik had tried to hold the spider, he would have succeeded. Arguably, this counterfactual is true. To put it in terms of the standard semantics for counterfactuals, it seems that the closest world where Erik even tries to hold the spider is one where his arachnophobia isn't as severe as it actually is. But then there isn't any obvious reason to think that Erik would have failed to hold the spider in that world – once we remove his crippling fear, it seems

[5] For discussion of this view, see Moore (1912), Nowell-Smith (1960), and Vihvelin (2013).
[6] This case is inspired by Keith Lehrer (1968). Instead of a fear of red candy, though, I use a fear of spiders.

perfectly within Erik's power to hold the spider. And if that counterfactual is true, then the Simple Conditional Analysis implies that Erik is free to hold the spider. That doesn't seem right.

In my view, cases like **Spider** aren't just bugs for the Simple Conditional Analysis to work out. Rather, such cases bring out a more general point about counterfactual analyses of the freedom to do otherwise. When evaluating counterfactuals, we are asked to consider the closest world where the antecedent is true (roughly, and according to the standard semantics). But we need to know whether the closest world where the antecedent is true *is close enough* to be relevant to the agent's freedom. In Erik's case, it may very well be that the closest world where he tries to hold the spider is one where he succeeds. But given that such a world requires a somewhat drastic change in Erik's psychology, that world is simply not close enough to be relevant to his freedom. We need some way of demarcating worlds that are close enough for freedom from worlds that aren't.[7]

FI provides at least a partial way of demarcating such worlds. According to FI, any world where the explanatorily independent facts are different is a world too far away to be relevant to freedom. More precisely: if actual fact *F* is distinct from and explanatorily independent of the facts constituting *S*'s behavior at time *t*, and *F* does not obtain in world *w*, then *w* is too far away to be relevant to *S*'s being free to do otherwise at time *t*. And this fits well with Erik's case. Plausibly, Erik's

[7] See Lehrer (1976), Fischer (1994, ch. 5; forthcoming), and Sartorio (2016; 2018) for discussions of this theme, albeit in somewhat different contexts.

crippling fear of spiders is in no way explained by his decision to run away and burst into tears at the sight of the spider. (If anything, the explanation goes the other way.) Hence, any world where he does not have such a crippling fear is irrelevant to his freedom. So, even if the closest world where he tries to hold the spider is one where he succeeds, FI implies that this world cannot be relevant to his freedom, at least insofar as Erik's psychology is different in this world.

So it may be that had Dave tried to keep his hand still, he would have, but this is not sufficient to show that Dave is free to keep his hand still. First, the Simple Conditional Analysis, which seems presupposed, is incorrect. Second, once we see the problem with the Simple Conditional Analysis, we see why principles like FI are all the more important: we need some way to demarcate worlds that are close enough to be relevant for freedom from worlds that are not close enough for freedom.

The Objection from the Metaphysics of Time

According to *eternalism*, all times and their occupants are equally real. Just as entities to the east and west of us are perfectly real, so eternalism states that entities before and after us are perfectly real. Of course, these (merely) past and future entities don't exist *now*, just as those entities to the east and west don't exist *here*. But they still *exist*. Contrasts to eternalism are views like *presentism*, which claims that only the present and its occupants are real: the past had its day, and the future will shortly, but only that which exists presently exists *period*.

Some authors might claim that, in order for FI to deliver intuitive results, we must presuppose the truth of eternalism and, hence, the falsity of views like presentism. Why? Let's focus on a particular case and then generalize. Suppose that at an earlier time, t_1, God believed that I would write this chapter now. Why did God have this belief? If eternalism is true, it's at least possible that my writing this chapter now explains God's past belief. That's because, under eternalism, when t_1 was present, the event of my writing this chapter existed in some sense. It didn't exist *at* t_1 of course, just as objects to the west don't exist *here*, but it still existed. Since God's belief and this event both exist, they can bear certain relations to each other, like an explanatory relation. And this explanatory relation obtaining is absolutely crucial for FI to deliver the result that God's past belief is not necessarily fixed for me now.

We have a much different story under presentism. If presentism is true, it seems incoherent to say that God's belief at t_1 is explained by my writing this chapter now. That's because under presentism, when t_1 was present, the event of my writing this chapter didn't exist *period*. Assuming that only existent entities can enter into relations, including explanatory ones, presentism implies that God's belief at t_1 was not explained by my writing this chapter when t_1 was present. Given the plausible supposition that what explains God's belief at t_1 doesn't vary over time, it follows that God's past belief is in no way explained by my current behavior. So, if presentism is true, then God's belief is fixed for me according to FI.[8]

Now this is just a particular case, and an admittedly controversial one, but we can generalize the problem to get around the controversy. At just an intuitive level, there seem to be many facts about the past and future that are not fixed for us now. We can disagree over individual examples, but most of us would agree that there are such facts. Let *F* be one such fact and let *e* be an event that *F* is about. (In the example above, *F* would be the fact that, at t_1, God believed that I would write this chapter now, while *e* would be the event of God's so believing.) In order for FI to deliver the result that *F* is not necessarily fixed for us now, it must be that *F* is at least partly explained by the facts constituting our current behavior, which seems to require that a particular relation obtain between *e* and our current behavior. But if presentism is true, it's hard to see how this could be. When *e* exists, our behavior does not; when our behavior does, *e* does not. So if presentism is true, it looks like *F* is in no way explained by the facts constituting our current behavior and, thus, fixed for us according to FI. It looks like we must presuppose eternalism in order for FI to avoid such problematic results.

This objection raises a host of deep issues and fully addressing it would take us too far afield. But I do think there is a relatively straightforward answer to it, even if the details are a bit messy. In its simplest form, the answer can be put as a

[8] See Fitch & Rea (2008) for an argument like this, although their targets are so-called "Ockhamist" views rather than FI.

dilemma: either presentism can accommodate cross-temporal relations, such as explanatory relations, or it can't. If it can, then presentism can also accommodate FI; if it can't, then presentism is too implausible to take seriously. Either way, FI goes unscathed.

I take it that the first horn of this dilemma is the controversial one, so let's spend some time unpacking it. Let's suppose that presentism can make sense of cross-temporal relations. In my view, the best strategy for the presentist begins by distinguishing between "ordinary" claims and "fundamental" claims. The basic idea is that ordinary claims are claims we routinely consider, accept, reject, etc., when outside of the metaphysics or philosophy room. They are claims about the way the world is "on the surface." Fundamental claims, on the other hand, are claims about the way the world is "deep below" – they are claims about what "underlies" those claims on the surface.[9]

Here are a couple of quick examples to get at this distinction. Suppose a military scout, upon seeing what he believes are enemy soldiers packing up camp and moving east, runs and tells the general "The enemy army is moving east!" There are all kinds of discoveries we could make that would render this claim false. Perhaps after retracing the scout's steps, we discover the scout was a bit turned around and actually saw the army moving west. Or perhaps we discover that the soldiers he observed only constitute a sliver of the enemy's army, and the rest of

[9] I get the language of "ordinary" and "fundamental" from Sider (1999; 2011; 2013). Indeed, much of what follows is heavily inspired by what Sider says.

the army is moving the other direction. Or maybe we learn that what he thought were soldiers were just local farmers, having nothing to do with the enemy army. These are all intelligible ways in which we would come to doubt the scout's claim. But here is what would *not* falsify the scout's claim: after doing some mereology and philosophy of language, we discover that there are no such things as "armies" – armies turn out to be mere pluralities of individual soldiers – and so the locution "*the* enemy army" has no referent. This kind of discovery, I contend, just isn't the kind of thing that would be relevant to evaluating the scout's claim.

One more case. Suppose you and a friend are having a fierce debate over the color of a long-lost jacket of yours, you thinking it had a slight tint of green and your friend thinking it was just grey. Again, we can think of a myriad of discoveries that would help us adjudicate the matter. Perhaps you come to learn that one of the lights above your coat rack that has a tint of green, thereby giving white and grey objects the appearance of being green as well. Or perhaps your friend comes to learn that he is slightly color-blind, and that he commonly mistakes green objects for grey ones. Or maybe the jacket slowly lost its green color over time, turning into grey eventually, and you and your friend are simply remembering different stages of the jacket. All of these discoveries would be helpful. But here is what would *not* help: after doing some quantum physics, we learn that there are no colored objects. There are only colorless quarks, electrons, neutrinos and the like which, when put together in the right way, reflect light at certain wavelengths, which then results in our minds (mistakenly) projecting those colors onto objects.

This discovery just isn't the kind of thing that would be relevant to the dispute between you and your friend.

Why is this, though? Why are mereological findings (largely) irrelevant to evaluating the scout's claim, or micro-physical findings (largely) irrelevant to settling the debate between you and your friend? It is because claims like "The enemy army is moving east" and "The jacket was green" are *ordinary* claims about the world – they are claims about the world "on the surface." In contrast, claims like "There are no composite objects" or "Colors are mere projections produced by the mind" are meant to be *fundamental* claims about the world – they are claims about what "underlies" those ordinary claims. And in general, learning what the world is like at the deepest levels is not very helpful in determining what it is like on the surface.

While the distinction between ordinary and fundamental claims may be intuitive, making it more precise is admittedly a bit tricky. One issue is whether ordinary claims can be *literally* true if the fundamental claims which underlie them don't line up in the right way, e.g., whether "The enemy army is moving east" can be literally true even if, fundamentally, there are no armies. My own sympathies lie with "liberal" views that answer affirmatively; more "conservative" views answer negatively, typically holding that such ordinary claims are at best "quasi-true."[10] Another issue is over how to best understand the "underlies" relation and

[10] For a claim to be "quasi-true" means, roughly, that two parties disagree over the truth-value of the claim but not because of any empirical disagreement, only a philosophical one. See Sider (1999; 2011) for elaboration.

the talk of "levels" of reality. Some authors talk about the fundamental claims "grounding" the ordinary ones, taking the picture of "levels" very seriously: although there is a certain priority to the more fundamental levels, each level is nonetheless equally real. Other authors prefer to talk about the fundamental claims as "carving nature at the joints," and thereby providing a "metaphysical semantics" for the ordinary claims. According to such views, there aren't different levels of reality, only levels of language, with the most fundamental carving nature perfectly at the joints.[11]

Fortunately, we don't need to enter into these disputes. So long as the distinction between ordinary and fundamental claims is accepted, the presentist need only defend two claims: (i) that presentism is a fundamental claim about the temporal nature of reality, and (ii) that claims involving cross-temporal relations, like the explanatory ones required for FI, are ordinary claims about reality. If both (i) and (ii) hold, then presentism can hope to accommodate claims about cross-temporal relations and, thus, also be squared with FI.

Let's work through some of the details here. Consider the following claim:

1. Socrates was a philosopher.

Initially, it might seem puzzling how the presentist could accept (1). After all, Socrates no longer exists and, hence, doesn't exist *simpliciter* according to

[11] For a defense of the former view, see Schaffer (2009); for a defense of the latter, see Sider (2011).

presentism. How then could (1) be true? Various responses have been given, but they all have significant drawbacks. For instance, some suggest that (1) expresses a proposition involving Socrates's *haecceity* – a property that Socrates alone can possibly exemplify – rather than Socrates himself. Others suggest that, although (1) seems to express a singular proposition, it really expresses a past-tensed existential proposition. Some instead opt for a Meinongian-style view according to which non-existent entities can still enter into relations. And more simply admit that (1) doesn't express any proposition at all and, hence, is not true, but that it nonetheless has "linguistic meaning" which might very well be true.[12] Each of these proposals has obvious costs.

Instead, I think the presentist should happily admit that (1) is true (or at least "quasi-true") *if* understood as an ordinary claim. Sure enough, presentism implies that Socrates doesn't exist *at the fundamental level*. But, so the presentist can say, this is not immediately relevant to evaluating (1) when understood as an ordinary claim. Just as the non-existence of armies *at the fundamental level* is irrelevant to the claim that "The enemy army is moving east," so the presentist should say that the non-existence of Socrates *at the fundamental level* is irrelevant to the truth of (1). If so, then the truth of (1) as an ordinary claim is not an immediate threat to presentism.

[12] For the first approach, see Adams (1986); for the second, see Bourne (2006); for the third, see Hinchliff (1988); for the fourth, see Markosian (2004).

Does this really help the presentist all that much, though? I believe so, and for two reasons. First, the presentist can claim that (1) only appears undeniable when understood as an ordinary claim. Indeed, it is unclear whether there are *any* undeniable fundamental claims (apart from mere tautologies perhaps). Consider the "multi-level" view of fundamentality again. Is it so obvious that the "bottom" level contains truths like (1)? Hardly. If we've learned anything from the "hard" sciences, it is that the lower levels are often strange and counterintuitive places. (Just think of the strangeness of quantum mechanics, or the conceptual revolution of spacetime.) It is simply unclear whether a claim like (1) is true at the fundamental level. And the same goes for "single level" views: it is unclear that the notions involved in (1) are perfectly joint-carving. So, while denying the truth of (1) understood as an *ordinary* claim seems quite problematic, denying its truth understood as a *fundamental* claim is not obviously problematic.

Second, and relatedly, it is conceptually possible for fundamental claims to look very different than the claims they underlie. Think of the scout's claim again, that "The enemy army is moving east." The fundamental claim which underlies this may very well be extremely complex in nature, at least if armies don't exist fundamentally. Likewise, the presentist should claim that the fundamental fact that underlies (1) looks quite different than what we might expect.

It may be helpful to have a somewhat fleshed-out account of how this might go. I'll explicate a version inspired by Ned Markosian's (2004) work, but there are certainly other (and perhaps even better) versions available.[13] First, the presentist can introduce a fundamental and irreducible modal propositional operator, 'P,' which intuitively stands for "It was the case that…" The presentist could then suggest something like the following claim underlies (1):

1a. P(∃x)(x is the referent of 'Socrates' and x is a philosopher).

Admittedly, (1a) looks quite a bit different than (1): (1a) involves an existential claim while (1) involves a singular claim; (1a) also involves a meta-linguistic claim whereas (1) does not. But if claims involving past individuals are merely ordinary claims, it's not surprising that the fundamental claims beneath them look quite different, or so the presentist should insist.

Now consider a claim involving a cross-temporal relation:

2. I admire Socrates.

Just as the presentist offered (1a) as the fundamental fact that underlies (1), so they can offer something like the following as the fundamental fact which underlies (2):

2a. There are various properties, p_1-p_n, such that I associate p_1-p_n with the name 'Socrates'; p_1-p_n evoke feelings of admiration in me; and P(∃x)(x is the referent of 'Socrates').

[13] Examples (1), (1a), (2), and (2a) are taken from Markosian. However, Markosian does not endorse the "underlying" strategy as stated. Instead, he thinks (1) and (2) do not, strictly speaking, express any proposition. A benefit of the "underlying" strategy is that it can allow for (1) and (2) to express propositions, albeit non-fundamental ones.

Here's perhaps an interesting point. By accepting that claims like (2a) underlie facts like (2), the presentist admits that there are no cross-temporal relations *fundamentally* – that the "bottom" level of reality or the perfectly "joint-carving" language contains no cross-temporal relations whatsoever. So there is some truth in saying that presentism cannot allow for cross-temporal relations. But claiming that there are no cross-temporal relations *fundamentally* seems far less problematic than saying there are no such relations *period*.

Let's now turn to a present event *causing* (or explaining) a future event under this suggestion. Consider the following claim:

3. My writing this chapter right now will (partly) cause you to read it later.

To get the fact which underlies (3), we introduce another fundamental and modal propositional operator, 'F,' which intuitively means "It will be the case that…" The claim which underlies (3) could look something like this:

3a. [F(∃x)(x is of the event-type of your reading this chapter)] partly because [(∃x)(x is of the event-type of my writing this chapter)].

A couple of notes about (3a). First, it is unlikely (to my mind) that either the event-type of "my writing this chapter" or "your reading this chapter" is fundamental. More likely, there is some very complex claim which underlies these as well. I appeal to them only for brevity. Second, the presentist should say more about the "partly because" relation if they can. Is it the same relation that obtains in cases of present, simultaneous causation? Does it have a distinctive inferential role? Does it only take facts as relata, or can it take events too? These are important questions, but I will sidestep them as my goal is only to sketch a way for the presentist to make sense of cross-temporal relations, including a causal one.

Finally, we can return to a case of a present event causing (or explaining) a past event, like our current behavior causing God's past beliefs. Consider the claim we started this section with:

4. God believed that I would write this chapter now because I am writing this chapter now.

The fact which underlies (4) could run as follows:

4a. [P(God believes: F(∃x)(x is of the event-type "my writing this chapter"))] partly because [(∃x)(x is of the event-type "my writing this chapter")].

Here is the critical point: (4a) seems perfectly consistent with presentism. So, if (4a) "underlies" claims like (4), then it seems as if we can make sense of God's past beliefs being explained by our current behavior, even under presentism. And if we can, then according to FI, divine foreknowledge is no threat to freedom.

So it seems to me that the best chance presentism has for accommodating claims involving cross-temporal relations is to insist that such claims are merely ordinary claims, and that the claims which underlie them do not invoke cross-temporal relations. But if this strategy is successful, then there is no reason to think that presentism raises trouble for FI: the truth (or "quasi-truth") of ordinary claims involving cross-temporal relations is enough for FI. That is, the first horn of the dilemma holds: if presentism can accommodate cross-temporal relations, it is through some version of the "underlying" strategy, which means presentism can also accommodate FI.

Now, at the end of the day, it may be that not even the "underlying" strategy can save presentism from the problem of cross-temporal relations. For one, each version of the "underlying" strategy, including the one I just presented, faces serious problems of its own. For another, some of our best physical theories seem to require *fundamental* cross-temporal relations. But this leads us to the second horn of the dilemma: if presentism cannot accommodate claims involving cross-temporal relations, then presentism is simply too implausible to take seriously. If presentism cannot account for the truth (or "quasi-truth") of claims like (1) through (4) even understood merely as ordinary claims, or if presentism is straightforwardly inconsistent with our best physical theories, then those sympathetic to FI should not lose any sleep over the relationship between FI and presentism.

As you can see, the details get a bit messy, and cleaning up the mess would take far too much space. But again, the basic response is fairly simple: either presentism can accommodate cross-temporal relations or it can't. Either way, FI is not in any trouble.

The Objection from Practical Reasoning

If my arguments from the last chapter are on target, then FI is quite appealing from a more theoretical perspective: FI provides an ecumenical and unified explanation for many distinct phenomena. But is FI all that appealing from a more "practical" perspective? That is, do we really invoke anything like FI in our practical reasoning? Or, to put it another way, does FI give us an intuitive "picture" to help us navigate the world?

For instance, think of FP again: that agent *S* can perform action *X* only if *S*'s performing *X* is consistent with the actual past. FP gives quite an intuitive picture of the world, namely, what some have called the "garden of forking paths." According to this picture, the past is a "single path" leading up to the present moment, while the future involves a series of "forking paths." Practically speaking, this means there is no point deliberating about the past – there is only one option! But the present and future are plainly worth deliberating about – think about how your present decisions will affect which "path" you end up on. Is there any such intuitive picture associated with FI?[14]

To make this challenge clearer, consider the following case taken from John Martin Fischer:

> **Icy Patch**: Sam saw a boy slip and fall on an icy patch on Sam's sidewalk on Monday. The boy was seriously injured, and this

[14] Thank you to John Martin Fischer for bringing these questions to my attention.

> disturbed Sam deeply. On Tuesday, Sam must decide whether to go ice-skating. Suppose that Sam's character is such that if he were to decide to go ice-skating at noon on Tuesday, then the boy would not have slipped and hurt himself on Monday. (1994, p.95)

Now imagine that Sam is trying to decide whether or not to go ice-skating. The following reasoning seems deeply suspect: "Well, if I were to go ice-skating, the accident wouldn't have occurred. That would be great for the boy who slipped. So, I should go ice-skating." As Fischer puts it, this kind of reasoning seems like "wishful thinking" at best.

FP and the garden of forking paths can explain why this line of thinking is so problematic. According to FP, we only have the power to add to the given past – there are no (longer) "forking paths" behind the present moment. Hence, deciding to perform some action based on a different past is simply confused. Cases like these show that FP presents an intuitive picture of the world, one that lines up with much of our practical reasoning.[15]

What about FI, though? Does FI have an intuitive picture, one that lines up with our practical reasoning? I offer two points in response. First, FI (or FI+) is perfectly compatible with the garden of forking paths, or at least an interpretation of it, and thus can also explain our intuitions in cases like **Icy Patch**. Second, there

[15] See also Fischer & Pendergraft (2013).

is an additional intuitive picture (or at least metaphor) associated with FI, one that is not brought out by principles like FP. Let's take each point in turn.

First, the garden of forking paths and FI (or FI+). When we think of the "single path" that leads up to the present moment, it is natural to think of that single path as representing the *external past* – that the single path contains all of the intrinsic facts that obtain *temporally prior* to the current moment. The "forking paths" beyond the present moment then represent all of the possible *temporal futures* consistent with the external past. But FI (or FI+) suggests an alternative interpretation. Instead of the single path representing the external past, it ought to represent the *causal (or explanatory) past* – that the single path contains all of the intrinsic facts that obtain *causally prior* to the current moment. The forking paths beyond the present moment then represent all of the possible *causal futures* consistent with the causal past. Of course, for us individuals who don't have access to time machines or the like, these two interpretations may very well come to the same thing. But the important point is that the advocate of FI (or FI+) need not give up on the deeply intuitive picture of the garden of forking paths. She need only offer a (slightly) different interpretation of it.

With this alternative interpretation in hand, FI (or FI+) can thereby accommodate our intuitions about cases like **Icy Patch**. Arguably, the boy's accident partly influences (at least ancestrally) Sam's decision, and so belongs on the single path leading up to Sam's decision. At the very least, the boy's accident is in no way caused by Sam's decision and so is on every one of the branching paths available to Sam. That is, FI (or FI+) implies that the accident is fixed for Sam at the time of his decision. But if the accident is fixed for him, it seems problematic for him to decide to go ice-skating based on the accident's non-occurrence. So FI also implies that Sam's line of thinking is confused.

On to the second point: not only is FI (or FI+) compatible with the garden of forking paths, but it also has an additional intuitive picture associated it, one not associated with FP. Consider the phrase that something is "out of one's hands." (For example, "I just submitted my job application; whether I hear from them is now out of my hands.") The picture (or metaphor) here is suggestive: when something is literally "out of one's hands," it means that one no longer has a grip on the object and, hence, does not have any (significant) influence over it, us being the tactile creatures we are. But what one lacks influence over, one also lacks control over. Hence, it is no wonder that when someone says an event is "out of her hands," she is implying that she no longer has control over the event.

This all fits quite nicely with FI. According to FI, if an agent has no explanatory influence over an event, then the event is beyond the agent's control or "fixed" for the agent. Indeed, in the last chapter I even phrased FI in terms of what is beyond an agent's causal or explanatory *reach*. So the picture (or metaphor) behind FI is this: given our circumstances, there are certain things in our hands or at least within our reach, while there are others that are out of our hands and beyond our reach. We presume ourselves to have some control over those things in our hands or within our reach; however, we lack control over those things out of our hands and beyond our reach. FI is a natural way of capturing and precisifying this picture. In contrast, FP does not fit well with this picture. There are many events in the present and the future that are "beyond our reach" and, hence, beyond our control. FP, by focusing solely on the past, does nothing to illuminate this metaphor.[16]

So FI (or FI+) sits nicely with both the garden of forking paths as well as the idea of something being out of one's hands, these pictures being ones we use quite routinely. This gives us good reason to think of FI (or FI+) as a foundational principle, one which undergirds much of our practical reasoning. But in closing this section, I'd like to note one last point, namely, that FI (or FI+) also fits with various *possible* but *non-actual* pictures of practical deliberation. For instance, in many science fiction novels, shows, and movies where time travel is feasible, there is often talk of "alternate timelines." Characters in such works are often said to have the ability to "jump to" or even "add" alternate timelines by time traveling. Of course, for creatures like us, these ideas have no import for our everyday lives. Moreover, it is somewhat difficult, philosophically speaking, to get a handle on what is behind such pictures, particularly whether they allow for a *changing* past

[16] Indeed, there are variations on this saying. Consider when someone says that an opportunity "slipped through her fingers," implying that she could have done something about it but no longer has the chance. Again, this phrase brings forth the picture of someone no longer having a grip on something and, hence, losing any ability to influence it.

or just *different* pasts.[17] But regardless, for creatures in various science fiction scenarios, these are popular and powerful pictures.

What's noteworthy is that FI can *explain* these kinds of pictures as well. In cases with time travel, the past is *not* beyond the agent's explanatory reach. Hence, FI implies that the external past doesn't necessarily comprise a single "path" for time-traveling agents; instead, it is a more complicated picture with many different "paths" or "timelines" preceding the present moment, the time-traveling agents having a choice as to which "path" or "timeline" to visit. Not only, then, does FI (or FI+) fit with actual pictures of practical deliberation, but it also sheds light on non-actual pictures as well.

The Objection from Prepunishment

Both Patrick Todd (2013) as well as Patrick Todd and John Martin Fischer (2013) provide a compelling challenge to FI (and argument for FP), one centered around *divine prepunishment*. I'll focus on Todd's (2013) paper, "Prepunishment and Explanatory Dependence," as the challenge is a bit more developed there, but what I have to say applies just as well to Todd's and Fischer's article. Todd writes:

> Suppose that ten days ago God prepunished Jones for sitting at *t*. And suppose Jones's punishment took the following form: spending ten hours in his local jail. So ten days ago Jones spent ten hours in

[17] See Law (2019) for discussion of this issue.

> his local jail. And he was punished by God in this way because he will sit at *t*. (p. 624)

At time *t,* is Jones free to refrain from sitting? Todd says it is fairly clear that Jones is not. But FI doesn't deliver this result. Jones's being in jail ten days ago seems to be partly explained by his sitting at *t* and, hence, his being in jail is not necessarily fixed for him at *t* according to FI. And if it is not necessarily fixed, then there is no reason to think that Jones must sit at *t*. By comparison, notice that Jones's being in jail ten days ago is plainly part of the past relative to *t* and, thus, fixed for him according to principles like FP. So, given that his being in jail entails his sitting at *t*, FP seems to get the right result that Jones is not free.

Todd puts the point more generally like this:

> In whatever sense it might be true that whether Jones spent ten hours in jail ten days ago 'depends on' whether he sits at *t*, this sense is *obviously irrelevant* to the question of what is within Jones's control at *t*. (p.629, emphasis in text.)

If Todd is right, then FI is just about doomed. After all, the sense in which Jones's jailtime depends on Jones's sitting at *t* seems to be the same sense that FI is concerned with. If this sense of dependence is irrelevant to freedom, then the central idea of FI – that those parts of the world which depend on our behavior are not necessarily threats to our freedom – is simply confused.

So is FI doomed? Others have tried to meet Todd's challenge,[18] but I believe the reflections of the last chapter can help meet the challenge in a superior way. In particular, recall that after considering various cases of time travel, I argued that we ought to augment FI as follows:

> **FI+**: Agent *S* can perform action *X* at time *t* (in world *w*) only if there is a world, *w'*, such that (i) all of the facts that are distinct from and explanatorily independent of the facts constituting *S*'s behavior at time *t* hold, (ii) all of the facts that *S*'s behavior at time *t* explanatorily depend on also hold, and (iii) *S* performs *X* at *t*.

The only difference between FI and FI+ is the addition of clause (ii) which claims, intuitively, that any fact which *explains* the agent's behavior at *t* is fixed for the agent at *t*. I argued that, in addition to being a plausible and minimal extension of FI, clause (ii) also allows us to treat certain cases of time travel in an intuitive way: FI+ implies that time travelers caught in explanatory loops (like Bill) aren't free, but does not imply as much for time travelers who aren't caught in explanatory loops (like Ted).

Once we shift from FI to FI+, Todd's challenge faces a dilemma: either prepunishment cases involve explanatory loops or they don't. If they do, then FI+ can accommodate the intuition that the prepunished individual isn't free. If they don't involve explanatory loops, it is not so clear that the individual's freedom is

[18] See Swenson (2016) for alternative responses.

undermined to begin with. Either way, prepunishment cases do not present a significant challenge to FI+. Let's walk through the argument .

Start with a couple of quick warmup cases. Consider the following: ten days ago, knowing that Jones will sit at *t*, God was a bit angry and needed to blow off some steam. So, ten days ago, God decided to cause a small explosion on a barren planet, galaxies away, but then immediately removed all traces of the explosion, ensuring that the explosion won't affect Jones at *t* whatsoever. In this case, is it obvious that Jones *must* sit at *t*? No. Indeed, given the obvious parallels between this case and Ted's not-so-excellent adventure from the previous chapter, it would seem that we should treat them the same. Since Ted's freedom isn't obviously undermined, so neither is Jones's.

Now a warmup case a little closer to home: ten days ago, knowing that Jones will sit at *t*, God wanted to "prepunish" *someone,* just not Jones. So, ten days ago, God decided to put Smith in jail for a very brief moment, but then immediately removed all traces of her jailtime, ensuring that it won't affect Jones at *t* whatsoever. (Perhaps God briefly put Smith in a jail on a distant and barren planet and then covered it up.) If Jones's freedom isn't obviously undermined by God causing a small explosion on a distant planet, why would Jones's freedom be undermined in this case? There would seem to be no principled difference between this warmup case and the previous one.

And finally, back to our original example: ten days ago, knowing that Jones will sit at *t*, God prepunished Jones by putting him in jail. Must Jones sit at *t*? Well let's suppose that there is no explanatory loop here – suppose God immediately removed all traces of Jones's jailtime, ensuring that it won't affect Jones at *t* whatsoever. (This might be a bit more difficult to imagine, a point to which we will return momentarily.) If Jones's freedom isn't obviously undermined in our warmup cases, it's again hard to see why it would be here. In all three cases, God brings about some event in response to foreknowing that Jones will sit at *t,* and then immediately removes all traces of that event, ensuring that the event in no way affects Jones at *t*. Why should it matter whether the event in question be an explosion, putting someone else in jail, or putting Jones himself in jail? Again, we need a principled reason for treating these cases differently. If we stipulate that there is no explanatory loop involved, it is quite difficult to see what this principled reason might be.

But what if we instead suppose that there is an explanatory loop? That Jones's jailtime does partly explain (even if only ancestrally) his sitting at *t*? For what it's worth, I suspect this supposition is forced on us given the details of Todd's original case. The idea is this: when God causes a small explosion or puts Smith in jail, these events need not affect the shape of Jones's past (relative to *t*), and so need not affect Jones's future, including his sitting at time *t*. But when God puts Jones in jail, that affects Jones's past (relative to *t*), and so *must* affect Jones future, including his sitting at *t*, however minute or indirect the effect may be. To use the language of four-dimensionalism, consider Jones's "time-slice" in jail ten days ago and his "time-slice" that is sitting at *t*. Jones's time-slice ten days ago comes *earlier* than his time-slice at *t,* and it is commonly held that an agent's time-slice, s_1, comes earlier than another one of her time-slices, s_2, only if s_1 causes (either directly or ancestrally) s_2 in the right way.[19] Hence, it follows that Jones's time-slice ten days ago partly causes (at least ancestrally) Jones's time-slice at *t*. By comparison, there isn't much reason to think that God's causing an explosion on a distant planet or punishing someone else *necessarily* affects Jones's time-slice at *t*.

Whether this is compelling or not is somewhat beside the point though. If we suppose that Jones's sitting at *t* is partly caused (at least ancestrally), and hence explained, by his being in jail ten days ago, FI+ delivers the result that Jones isn't free at *t*. Since Jones's jailtime partly explains (at least ancestrally) his sitting at *t*, clause (ii) of FI+ implies that it is fixed for him at the time of his sitting.[20] Given that his jailtime entails his sitting, Jones's freedom is undermined according to FI+.

To be clear, we need not suppose that Jones's jailtime ten days ago is explanatorily *salient* with regards to Jones's sitting. In a typical context, the most natural explanation for Jones's sitting at *t* will not cite his being in jail ten days ago. Rather, it will cite things like Jones's needing a rest, or being asked to sit, or what have you. We need only suppose that Jones's jailtime *partly* explains (at least ancestrally) his sitting at *t*: that his being in jail ten days ago explains why he was wearing an orange jumpsuit and staring at a brick wall, which in turn explains why

[19] See Sider (2001, ch.1) and Wasserman (2018, ch. 1).

[20] Again, I am assuming that explanation is a transitive notion for expository purposes.

he started to grow bored, which in turn explains why he… which in turn explains why he decided to sit at *t*. So long as we construe the notion of explanation in a broad way, we can make sense of Jones's jailtime explaining (at least ancestrally) his sitting at *t*, regardless of whether his jailtime is explanatorily *salient* or not.

The dilemma for prepunishment cases is now clear: either the prepunishment partly explains (at least ancestrally) the agent's later activity or it doesn't. If it does, then FI+ delivers the right result; if it doesn't, then it isn't clear that the agent's freedom is undermined to begin with. Either way, the shift from FI to FI+ makes for a new reply to the challenge of prepunishment.

The Circularity Objection

Perhaps the most troubling objection to FI is that it seems to be circular. Compare the following two cases:

> **Working Time Machine**: Tim has before him a working time machine: all he has to do is press the button and he'll be sent to the distant past. Tim decides time travel is too risky though, and refrains from pressing the button, never traveling to the past.
>
> **Non-working Time Machine**: Tim has before him what he *thinks* is a working time machine, but he is radically mistaken: if he presses the button, the machine will instantly blow up. Tim decides time

> travel is too risky though, and (fortunately) refrains from pressing the button, never traveling to the past.

Intuitively, the past is explained by Tim's behavior in the first case, but not the second. What's the difference? One quite plausible answer is that, in the first case, Tim is *free* to travel to the past. With a working time machine, Tim *can* (in the relevant sense) travel to the past; without one he can't. If this is the right explanation, then whether a fact is explained by an agent's behavior is determined, in part, by what the agent *can and can't do*. Combine this with FI and we get a vicious circle: FI claims that what an agent can and can't do is determined, in part, by which facts are explained by the agent's behavior. But whether a fact is explained by the agent's behavior is determined, in part, by what the agent can and can't do. This is circularity at its worst.[21]

Before responding to the circularity objection, it's worth noting that there is another, intimately related objection to FI. I'll call it the "explosion" objection, and it centers around how to best understand which facts constitute an agent's "current behavior." Focus on some ultra-mundane fact about the past, like the fact that 100 years ago, a certain rock wasn't knocked over. Here is a fact about my behavior: I am *not* traveling to the past to knock over the rock. This is a bizarre fact to be sure, especially since I don't have anything even remotely close to a time machine at my disposal. But it nonetheless seems to be a fact about my behavior, one that at least

[21] Thank you to Neal Tognazzini for this objection.

partly explains (in a non-salient way) why the rock wasn't knocked over – after all, if I were to travel to the past to knock over the rock, I might succeed. FI doesn't imply that the rock's position is fixed for me since its position is partly explained by facts constituting my current behavior.

We can get a similar result even without any mention of time travel. Focus instead on some ultra-mundane fact about the present, like the fact that there is a rock on Neptune that is not presently being knocked over. Again, there seems to be a (bizarre) fact about my behavior that at least partly explains this fact: that I am not teleporting to Neptune right now to knock the rock over. If so, then FI again fails to imply that the rock's position is fixed for me.

More generally, cases like these seem to show that FI hardly implies that *any* fact is fixed for the agent, at least given a fairly liberal view of what counts as a fact about the agent's current behavior. The number of facts that are dependent on the agent's current behavior "explodes" so much as to render FI completely uninformative.[22]

Now the most natural response to the explosion objection would be to restrict FI to only *some* facts of the agent's current behavior. A tempting way to do this would be to restrict FI to facts about what the agent doesn't do but *could have*. Even if it is true that I am not traveling to the past or to Neptune to knock over the rock in question, I *couldn't have* done so, given that I have neither a time machine

[22] Thank you to Ryan Wasserman for raising this objection.

nor a teleporter. But this reply leads right back into the circularity worry. If we restrict our attention to facts about the agent's behavior, facts that the agent *could have* made false, FI is again made circular: FI would be relying on claims about what an agent can and can't do to determine whether a fact is fixed for the agent or not.

These are powerful and deep objections to FI and, to be frank, I don't have fully satisfactory responses. But I do think there are some fairly promising suggestions, promising enough to assuage our worries. In what follows, I'll first propose a specific restriction on FI that seems to make some progress in answering both the circularity objection and the explosion objection. I'll then conclude by giving more general reasons for being suspicious of each objection, even should the proposed restriction fail.

To start, I grant that we should indeed restrict FI in a certain way, but I deny that the most plausible restriction renders FI circular. Most basically, the proposal is this: instead of looking at facts about what the agent *can or can't do,* as these objections suggest, we ought to look at facts about what the agent *tries or doesn't try to do*. Somewhat more precisely, we should reformulate FI as follows: if fact *F* is distinct from and in no way explained by any of the facts about what *S tries* or *doesn't try* to do at *t*, then *F* is fixed for *S* at *t*. By doing so, FI can make significant headway in answering both objections, or so I will argue now.[23]

Let's start with the explosion objection. It is certainly true that I am not traveling to the past or being teleported to Neptune, and these facts may very well explain why the rocks in question aren't knocked over. But what about the fact that I am not *trying* to travel to the past or to teleport to Neptune? Do these facts explain why the rocks aren't knocked over? Presumably not. Since I have neither a time machine nor a teleporter, if I had tried to travel to the past or to Neptune, I certainly would have failed. More generally, given my actual circumstances, there seems to be no action such that, if I were to *try* to perform that action, the rocks in question would (or might) be knocked over. That strongly suggests that no fact about what I am currently trying or not trying to do in any way explains why the rocks aren't knocked over: the fact that the rock in question isn't knocked over seems explanatorily independent of what I am currently *trying* or *not trying* to do, even if the fact is partly explained by what I am currently *doing* or *not doing*. More generally, FI restricted as such implies that there are plenty of facts that are fixed for us, thereby helping to diffuse the explosion objection.

Admittedly, there will be some tricky cases. For instance, suppose that I have a working time machine in front of me, but I don't know the keycode to activate it and so don't even bother trying to use it. Is the past dependent on my

[23] Interestingly, a somewhat similar issue and reply comes up with regards to Vihvelin's so-called "Counterfactual Principle" of freedom. See Vihvelin (1996), Vranas (2010), and especially Wasserman (2018, ch. 4).

not trying to use the time machine? And is the past fixed for me? At least in my own case, there is some ambivalence about both questions. On the one hand, I could have tried a random code and gotten massively lucky – it wasn't *impossible* for me to travel to the past, just incredibly unlikely. But on the other hand, I had no clue what the code was! If I had tried to travel to the past, I almost certainly would have failed.

Whatever we make of cases like these, I don't think they present significant problems for the restriction of FI under consideration. Since our intuitions are ambivalent here, such cases don't seem to pose a significant threat for FI. And moreover, there are some plausible ways of dealing with this ambivalence consonant with, if not brought out by, this restriction of FI.[24] So even if the proposed restriction doesn't *completely* diffuse the explosion objection, it at least helps to contain it.

Let's now turn to this restriction and the circularity objection. Recall the cases of **Working Time Machine** and **Non-working Time Machine**. We started by noting that, intuitively, only in the former case is the past explained by Tim's behavior. That still seems right. The next step of the objection was to suggest

[24] Briefly, here are two ways. First, one might think that, since it was possible for me to use the time machine, the past is not fixed for me, but that my (non-culpable) ignorance makes it so that I am not responsible for the past in any way. Our ambivalence results from failing to distinguish the "control" condition of freedom (or moral responsibility) from the "epistemic" condition of freedom (or moral responsibility). Second, one might think that our ambivalence is simply due to us being imprecise about what it means for me to try to use the time machine. If we are being fine-grained enough to where a relevant precisification of my "trying to use the time machine" includes my "trying to push the correct keycode," then perhaps the past isn't fixed for me. But if we are being more coarse-grained and don't count this as a relevant precisification, then perhaps the past is fixed. FI fits with both approaches.

that this difference is determined by the difference in what Tim *can* and *can't* do: that the past is explained by his behavior in the first case because he *can* travel to the past. It's this second step that my response challenges. Instead of accounting for the difference in explanatory structure by appealing to what Tim *can* or *can't* do, I'm suggesting we should (or at least can) account for the difference in terms of what Tim *tries* or *doesn't try* to do. In neither case does Tim even *try* to travel to the past, but only in the first case does this *explain* the past: only in the first case is it plausible that, if Tim had tried to travel to the past, the past would (or might) have been different.[25]

Now we might wonder whether the circularity objection can be rerun against this restricted version of FI. Initially, the objection (in conjunction with the explosion objection) points out that focusing merely on facts about what the agent *doesn't do* is far too broad, and that we should restrict our attention to facts about what the agent doesn't do *but could have*. My response is to concede that focusing on facts about what the agent doesn't do is too broad, but that we can avoid circularity by instead focusing on facts about what the agent doesn't *try* to do. But, so the thought might go, this still isn't enough: we need to restrict our attention to

[25] Importantly, I am not relying on anything like a counterfactual analysis of causation or explanation in answering these objections. Rather, I take counterfactual (in)dependence to merely serve as a *heuristic* for explanatory (in)dependence. This isn't too controversial, as it is widely agreed that there is some relation between explanatory (or at least causal) dependence and counterfactual dependence, even if it is hard to say what that relationship is exactly. See Schaffer (2005) and Sartorio (2013) for helpful remarks here.

facts about what the agent doesn't try to do *but could have*. If so, then the circularity objection would still get off the ground.

However, it seems to me that restricting FI in this way goes *too far*. There seem to be plenty of cases where an agent *can't even try* to perform the relevant action, but her behavior still seems explanatory. Consider one final case:

> **Controlled Connie**: Connie is tempted to kick a nearby rock over during her walk home. After deliberating a while, she resists the urge and so the rock remains still. Unbeknownst to Connie though, all of her thoughts and decisions, including her not even trying to kick the rock over, are completely controlled by nefarious neuroscientists.

Presumably, Connie can't even *try* to kick the rock over given the presence of the nefarious neuroscientists. But it still seems as if her behavior at least partly explains the rock's remaining still. Her not trying to kick the rock isn't anything like the *ultimate* explanation, and she arguably isn't *blameworthy* for not trying to kick the rock over. But all the same, if someone wanted to know why the rock remained still, it would seem perfectly legitimate to cite the fact that Connie didn't try to kick the rock over – after all, if she had tried to kick over the rock, she might have succeeded. The more general lesson is that in determining whether or not a certain fact is explained by an agent's behavior, we should *not* focus solely on facts about what the agent could have done. And if we don't, then the circularity objection cannot be rerun.

So it seems to me that restricting FI to the facts about what the agent tries and doesn't try to do rescues it from both the circularity objection and the explosion objection. However, I must admit that there is some (hopefully remote) possibility that even the restricted version of FI will face some version of these objections – philosophers are quite remarkable at coming up with counterexamples, after all. In concluding this section, then, I'd like to offer more general reasons to be skeptical of both objections.

Here is the general reason to doubt the circularity objection. In order for an analysis of notion *X* to be circular, it must be that the analysis uses some notion, *Y*, that is partly analyzed in terms of *X*.[26] Now, FI is a partial analysis of a particular sense of "can," the sense relevant to *freedom*,[27] and it appeals to the notions of "worlds," "facts," "behavior," "time," and "explanation." So, if FI is a circular analysis, it must be that at least one of these notions is then analyzed in terms of the particular sense of "can" or "freedom" we are interested in. Presumably, neither "worlds," "facts," "behavior," nor "time" is to be analyzed in terms of "freedom." Hence, it seems that if FI is a circular analysis, it is "explanation" that is to be analyzed in terms of "freedom." And, sure enough, that is what cases like **Working Time Machine** and **Non-working Time Machine** are meant to show: that

[26] As an amusing example, the 2007 edition of the Merriam-Webster dictionary defines a "hill" as "a usually rounded elevation of land lower than a mountain." But it then defines a "mountain" as "a landmass that projects conspicuously above its surroundings and is higher than a hill."

[27] Or better: there isn't a particular *sense* of "can" at issue in FI. Rather, an agent is *free* to do otherwise just in case the agent can do otherwise *under a certain restriction*. FI is meant to make part of that restriction explicit.

certain facts are explained by an agent's behavior only if the agent is free to perform certain actions.

Is this correct, though? Should we analyze "explanation" in terms of "freedom"? This seems wrongheaded to me. In a nutshell, here's why: if giving an analysis of "explanation" requires appealing to "freedom," then it must be that "freedom" is a more *basic* notion than "explanation." But that seems implausible. So, it must be that, however we analyze "explanation," we need not (and should not) appeal to "freedom." Allow me to elaborate.

I take it that giving a genuine *analysis* of a notion involves analyzing that notion using more *basic* notions. For instance, the analysis of "bachelor" is "an unmarried human male" because all of the notions in the latter are more basic, in some sense, than the notion of the former. It's not just that the latter provides necessary and jointly sufficient conditions for the former. After all, the former provides a necessary and sufficient condition for the conjunction of the latter as well. Rather, the latter is a proper analysis of the former because "unmarried," "human," and "male" are all more basic notions than "bachelor." Or, to use a more philosophical example, when David Lewis (1973) provided his "counterfactual dependence" analysis of causation, presumably he didn't just think he was providing necessary and jointly sufficient conditions for when it was true that "x causes y"; rather, he also thought that the notions involved in counterfactual dependence were more basic than that of causation.

Now, it's a vexed question as to what, exactly, makes one notion more "basic" than another, one we need not answer here. Instead, I merely claim the following: that the notion of "explanation" is more basic than the notion of "freedom." Or, at the very least, the notion of "freedom" is not *more* basic than the notion of "explanation." Why should we think this? It will help to focus on the paradigm case of "explanation," namely, "*causal* explanation." It seems eminently plausible that "causal explanation" is more basic than "freedom" or "free action." To see this, notice that many authors are inclined to analyze "action" partly in terms of "causal explanation." For instance, one influential analysis holds, roughly, that an agent performs an action just in case the causal explanation of the agent's behavior involves, in the right way, a belief-desire pair had by the agent. Another influential analysis holds, roughly, that an agent performs an action just in case the causal explanation of the agent's behavior involves, in the right way, an intention had by the agent.[28] If such analyses are on target, then it must be that "causal explanation" is more basic than "action" and, *a fortiori*, more basic than "free action."[29]

Granted, this only gives us reason to think that "causal explanation" is more basic than "freedom" or "free action"; it does not give us reason to think that *all* kinds of "explanation" are more basic than "freedom." But similar comments apply

[28] For the *locus classicus* of each view, see Davidson (1963) and Davidson (1978) respectively.
[29] It may be that "*S* is free to perform *X*" *implies* that "*X*'s occurrence (or non-occurrence) is partly explained by *S*'s behavior," but that does not show that the notions in the former are more basic than the notions in the latter. For instance, "*S* is a bachelor" implies that "*S* is unmarried," but the notions of the former aren't more basic than the notions of the latter; if anything, it's the opposite.

to other notions of explanation: it seems that neither "metaphysical explanation," nor "conceptual explanation," etc. need be analyzed in terms of "freedom." If anything, it ought to go the other way. So, at the very least, "freedom" isn't more basic than these notions.

If the foregoing remarks are on track, then the notion of "explanation" is not to be analyzed in terms of "freedom." Assuming that no other notion involved in FI is to be analyzed in such terms either, which seems quite plausible, it is dubious that FI suffers from any kind of problematic circularity.

In regards to the explosion objection, the general reason for doubt is less precise, but perhaps more intuitive. Think about our rock cases again. I think all parties will grant that there at least *seems* to be an important causal (and hence explanatory) difference between my not kicking the rock on Neptune and Connie's not kicking the rock on her walk home. Any view which said otherwise would have some explaining to do. Perhaps I'm mistaken and the difference isn't best captured by facts about what the agent tries or doesn't try to do. But so long as there is *some* difference, as it surely seems there is, FI can be salvaged.

It's also important to note that it is not just FI which assumes an answer to something like the explosion objection. For instance, consider FP once again: that the past is fixed for us now. As noted in the last chapter, authors often admit that FP presupposes a lack of backward causation.[30] But if the explosion objection is

[30] See Fischer & Todd (2011) and Sartorio (2015).

taken seriously, then backwards causation is rampant: for just about any given past fact, it obtains partly because we are not now traveling to the past to prevent it from obtaining. Since FP is such an intuitive principle, I'm inclined to think that this gives us all the more reason to not take the explosion objection very seriously, even if it is hard to say where it goes wrong. At the very least, the explosion objection isn't *unique* to FI; it is only more *obvious* under FI.

All of this to say, even if the particular restriction of FI I've proposed fails, there are general reasons to doubt both the circularity objection and the explosion objection. As with most problems in philosophy, the devil is in the details. But hopefully I've convinced you that we have enough here to be skeptical of both objections.

Concluding Remarks

I have considered five objections to FI: the Symmetric Universe objection, the objection from the metaphysics of time, the objection from practical reasoning, the objection from prepunishment, and the objection from circularity. I've contended that, while all objections help clarify FI, none of them, except for possibly the last, pose significant trouble. To be sure, there are other objections to FI, but they will have to wait for another time. Instead, I will assume that we now have reason to take FI seriously. With that in mind, I will spend the next chapter exploring FI's implications for freedom.

References:

Adams, Robert M. (1986). "Time and Thisness," *Midwest Studies in Philosophy* XI: 315-329.

Bourne, Craig (2006). *A Future for Presentism*. Oxford: Oxford University Press.

Davidson, Donald (1963). "Actions, Reasons, and Causes," *Journal of Philosophy* 60: 685-700.

Davidson, Donald (1978). "Intending," *Philosophy and History of Action* 11: 41-60.

Fischer, John M. (1994). *The Metaphysics of Free Will*: *An Essay on Control*. Malden: Blackwell Publishing.

Fischer, John M. (forthcoming). "Local-Miracle Compatibilism: A Critique," in *Free Will: Historical and Analytical Perspectives*, edited by Jörg Noller and Marco Hauser. London: Palgrave/Macmillan Press.

Fischer, John M. & Pendergraft, Garrett (2013). "Does the Consequence Argument Beg the Question?" *Philosophical Studies* 166: 575-595.

Fischer, John M. & Todd, Patrick (2011). "The Truth about Freedom: A Reply to Merricks," *Philosophical Review* 120: 97-115.

Fitch, Alicia & Rea, Michael (2008). "Presentism and Ockham's Way Out," *Australasian Journal of Philosophy* 84: 511-24. Reprinted in *Freedom, Fatalism, and Foreknowledge* (2015), edited by John M. Fischer and Patrick Todd. Oxford: Oxford University Press: 229-246.

Hinchliff, Mark (1988). *A Defense of Presentism*. Ph.D. Dissertation, Princeton University.

Law, Andrew (2019). “The Puzzle of Hyper-Change,” *Ratio* 32: 1-11.

Lehrer, Keith (1968). “Cans without Ifs,” *Analysis* 29: 29-32.

Lehrer, Keith (1976). “‘Can’ in Theory and Practice,” in *Action Theory*, edited by Myles Brand and Douglas Walton. Dordrecht: Reidel: 241-270.

Lewis, David (1973). “Causation,” *Journal of Philosophy* 70: 556-567.

Locke, John (1690/1975). *An Essay Concerning Human Understanding*. Oxford: Oxford University Press.

Markosian, Ned (2004). “A Defense of Presentism,” in *Oxford Studies in Metaphysics,* vol. 1, edited by Dean Zimmerman. Oxford: Oxford University Press: 47-82.

Moore, George E. (1912/2006). *Ethics*. Oxford: Oxford University Press.

Nowell-Smith, Patrick H. (1960). “Ifs and Cans,” *Theoria* 26: 85-101.

Sartorio, Carolina (2013). “Making a Difference in a Deterministic World,” *Philosophical Review* 122: 189-214.

Sartorio, Carolina (2015). “The Problem of Determinism and Free Will Is Not the Problem of Determinism and Free Will,” in *Surrounding Free Will*: *Philosophy, Psychology, and Neuroscience*, edited by Alfred Mele. Oxford: Oxford University Press: 255-273.

Sartorio, Carolina (2016). *Causation and Free Will*. Oxford: Oxford University Press.

Sartorio, Carolina (2018). “Situations and Responsiveness to Reasons,” *Noûs* 52: 796-807.

Schaffer, Jonathan (2005). “Contrastive Causation,” *Philosophical Review* 114: 327-358.

Schaffer, Jonathan (2009). “On What Grounds What,” in *Metametaphysics: New Essays on the Foundations of Ontology*, edited by David Chalmers, David Manley, and Ryan Wasserman. Oxford: Oxford University Press: 347-383.

Sider, Theodore (1999). “Presentism and Ontological Commitment,” *Journal of Philosophy* 96: 325-347.

Sider, Theodore (2001). *Four-Dimensionalism: an Ontology of Persistence and Time*. Oxford: Oxford University Press.

Sider, Theodore (2011). *Writing the Book of the World*. Oxford: Oxford University Press.

Sider, Theodore (2013). “The Evil of Death: What Can Metaphysics Contribute?” in *The Oxford Handbook of Philosophy of Death*, edited by Ben Bradley, Fred Feldman, and Jens Johansson. Oxford: Oxford University Press: 155-166.

Todd, Patrick (2013). “Prepunishment and Explanatory Dependence: A New Argument for Incompatibilism About Foreknowledge and Freedom,” in *Philosophical Review* 122: 619-639.

Todd, Patrick & Fischer, John M. (2013). “The Truth about Foreknowledge,” *Faith and Philosophy* 30: 286-301.

van Inwagen, Peter (1983). *An Essay on Free Will*. Oxford: Oxford University Press.

Vihvelin, Kadri (1996). “What Time Travelers Cannot Do,” *Philosophical Studies* 81: 315-330.

Vihvelin, Kadri (2013). *Causes, Laws, and Free Will: Why Determinism Doesn’t Matter*, Oxford: Oxford University Press.

Vranas, Peter (2010). “What Time Travelers May be Able to Do,” *Philosophical Studies* 150: 115-121.

Wasserman, Ryan (2018). *Paradoxes of Time Travel*. Oxford: Oxford University Press.

Chapter 3: Implications of the Fixity of the Independent

Introduction

The case for FI has been set. We now turn to its implications, in particular what it has to say about various incompatibility arguments. I'll focus on the big three: arguments for the incompatibility of the freedom to do otherwise and future contingents, divine foreknowledge, and physical determinism.[1] I will contend that FI renders standard arguments in the first two categories inert. In contrast, FI bolsters and sheds light on arguments in the third category.

Before proceeding, I wish to convey two points to the reader. First, many of the issues in this chapter are dialectical rather than logical. It's not too difficult to see that FI suggests that arguments involving future contingents and divine foreknowledge are uncompelling, but implies that determinism is incompatible with freedom (for agents like us). It is less clear whether those who disagree with these deliverances of FI should feel compelled to accept FI. Second, FI's relation to arguments involving divine foreknowledge has received the most attention so I spend the most time there.

[1] If the reader is interested in FI and its relation to more esoteric forms of determinism, see my "Free Will and Two Local Determinisms," with Neal A. Tognazzini.

FI and Future Contingents

Here's how a standard argument for the incompatibility of the freedom to do otherwise and future contingents goes: suppose that either it is true at time t_1 that *S* will perform *X* at time t_2 or it is false (where t_1 is earlier than t_2). If it is true at time t_1 and yet *S* could have done otherwise than perform *X*, then *S* could have performed an action that would have required the past to be different – necessarily, if *S* does otherwise than *X* at t_2, then a fact that actually obtains at t_1 does not obtain. Likewise, if it was false at time t_1 and yet *S* could have performed *X*, then *S* also could have performed an action that would have required the past to be different. But no agent can perform an action such that her performing it requires the past to be different – the past is "fixed." So, if it is either true or false at t_1 that *S* will perform *X* at time t_2, then *S* couldn't have done otherwise than what she in fact does at t_2. Let's call this the "argument from future contingents."

How does FI respond? Consider the premise that no agent can perform an action such that her performing it requires the past to be different. This premise just is a version of the principle of the fixity of the past (FP). But if my arguments of the first chapter are correct, then FP is only a derivative and approximately correct principle and instead should be replaced by FI. And FI implies, *contra* FP, that if a part of the past is partly explained by the agent's relevant behavior, then it is not necessarily "fixed." But that is exactly what we seem to have in the case of future contingents. Even if it is true (or false) at t_1 that *S* will perform *X* at t_2, it is arguably true (or false) *because* of what *S* does at t_2.[2] So, FI insists that this is a case where FP might lead us astray – this is a case where an agent might very well have the ability to perform an action that would require the past to be different.

That's how FI responds, but should those who are sympathetic with the argument from future contingents be moved at all? At the very least, should they be moved by my defense of FI? Here's one reason to think not: one of my arguments for FI involved the assumption that future contingents are *not* a threat to freedom. Here's how the argument went: those who accept a version of FP usually restrict it in such a way that future contingents are not necessarily fixed – it is usually restricted to the "temporally intrinsic" or "hard" past, which seems to exclude future contingents. What justifies this restriction? FI provides an answer: the temporally intrinsic hard past is (usually) not explained by any agent's current behavior whereas the "temporally extrinsic" or "soft past" sometimes is.

But if one is not inclined to restrict FP in such a way, then this argument has no momentum whatsoever. In fact, it is clear that this argument for FP plainly begs the question against authors who think that future contingents, if there are any,

[2] We might wonder whether future contingents are explained by the agent's future behavior if certain other metaphysical views are correct. For instance, Michael Rea (2006) argues that if *presentism* is true – the view, roughly, that only the present is real – then future contingents are not explained by the agent's future behavior. (Finch & Rea (2008) seem to raise a similar worry about divine foreknowledge.) In short, I don't find these arguments terribly compelling. Either presentism can accommodate cross-temporal relations, such as certain explanatory relations, or it can't. If it can, then FI will deliver the same results; if it can't, then presentism isn't worth taking seriously. Either way, FI is in no trouble. See the last chapter for discussion.

are fixed. If so, then this argument puts no pressure on those who accept the argument from future contingents.

I think it's fair to say that the argument for FI just mentioned should not move those who accept the argument from future contingents. But there are other arguments for FI. For instance, consider the very first argument I presented: that although the past seems fixed, the future does not, and that the best explanation of this asymmetry is that the past is beyond our explanatory reach and that FI is true. This argument does not obviously beg the question against those who accept the argument from future contingents. On the contrary, it could even grant an unrestricted version of FP. All that it depends on is the relevant asymmetry between past and future.

So, the first argument gives us good reason to accept FI, and in a way that doesn't seem to presuppose that FP ought to be restricted to the "temporally intrinsic" or "hard" past. Then there are also the other arguments: the argument from relativity and the argument from time travel. These arguments also seem forceful regardless of whether we restrict FP to the hard past. Assuming these arguments are compelling, it seems that there is still a solid case for FI independent of whether one thinks FP ought to be restricted to the hard past or not. Let's now move on to the case of FI and divine foreknowledge, a subject which has received far more attention.

FI and Divine Foreknowledge[3]

Here's how the standard argument for the incompatibility of the freedom to do otherwise and divine foreknowledge goes: suppose that at time t_1, God knows (and hence believes) that *S* will perform *X* at t_2 (where t_1 is earlier than t_2). If God knows this and yet *S* could have done otherwise than *X* at t_2, then either *S* could have made God have a false belief at t_1, or *S* could have performed an action that would have required the past to be different (namely, God's belief at t_1). But no agent can make God have a false belief, and no agent can perform an action that requires the past to be different. So, if God knew at t_1 that *S* will perform *X* at t_2, then *S* cannot do otherwise than *X* at t_2.[4] Let's call this the "argument from divine foreknowledge."

Again, it is clear how FI responds. A key premise in the argument is FP, but FP ought to be abandoned for FI: if a part of the past is partly explained by the agent's behavior, then it is not necessarily "fixed." And, arguably, that is what we have in the case of divine foreknowledge. Even if at t_1 God believes that *S* will perform *X* at t_2, God arguably believes this *because* of what *S* does at t_2: generally, God believes that *P* is true because *P* is true, not the other way around. Hence, God's past beliefs are not necessarily fixed. Following others, let's call this the "dependence" response.

[3] Much of what follows in this section comes from my paper "Freedom and Dependence: A Dialectical Intervention," with Taylor W. Cyr. Thank you to Dr. Cyr for allowing me to use that material here.

[4] See Pike (1965) for the classic presentation of this argument.

The dependence response has been around in one way or another since at least Merricks (2009), perhaps all the way back to Origen (246/2002). Unfortunately though, it has not always been explicated in a careful way. As a result, the dependence response has been charged with being dialectically infelicitous. The first objection charges the dependence response with begging the question against those who endorse the argument from divine foreknowledge. The second objection charges the dependence response with failing to advance the dialectic in a helpful way since it is not a new position, but rather a version of some well-worn response. We'll take each objection in turn.

The Charge of Begging the Question

John Martin Fischer and Neal Tognazzini (2014) have argued that several recent presentations of the dependence response – defended by Trenton Merricks (2009; 2011), Storrs McCall (2011), and Jonathan Westphal (2011; 2012) – "fail to respect the relevant dialectical context by presupposing the very thing that is called into question by the [argument from divine foreknowledge]." (2014, p. 362) In particular, Fischer and Tognazzini claim that these authors' responses to the argument rely on the claim that we have a choice about what we do (and thus what God believed we would do), which is to say that we have the freedom to do otherwise than what God believed we would do. But this claim, Fischer and Tognazzini argue, is dialectically unavailable to the proponents of the dependence response because it is precisely this claim that is called into question by the argument from divine foreknowledge. Thus, Fischer and Tognazzini conclude, the dependence response simply begs the question.

To be fair to Fischer and Tognazzini, it might very well be that the dependence response *as presented by these other authors* does beg the question. But I don't think my presentation does. Here's how I see the dialectic: first, an argument is presented, one that aims to call into question the intuitive idea that we possess the freedom to do otherwise, at least supposing God has exhaustive foreknowledge of what we do. According to the dependence response, however, the mere addition of God's past beliefs does not call into question our freedom to do otherwise. By rejecting FP in favor of FI, the rationale is this: given that (some of) God's past beliefs explanatorily depend on what we now do, they are not necessarily fixed for us now. And if they aren't necessarily fixed, then divine foreknowledge *itself* doesn't call into question our freedom to do otherwise.

Crucially, the proponent of the dependence response is not simply insisting that we do have the freedom to do otherwise (regardless of whether God has exhaustive foreknowledge). Rather, the proponent of the dependence response is claiming that if we could have done otherwise *absent* God's past beliefs, then the mere addition of God's past beliefs does not undermine our freedom to do otherwise. *Contra* Fischer and Tognazzini, nothing about this response requires a commitment to the claim that we do in fact have the freedom to do otherwise. Or to put it another way, it is perfectly coherent to accept the dependence response and yet, for independent reasons, deny that we have the freedom to do otherwise.

If this is how the dialectic goes, it is clear that the dependence response does not necessarily beg the question against those who accept the argument from divine foreknowledge.

The Charge of Failing to Advance the Dialectic

So, the dependence response does not beg the question against those who accept the argument from divine foreknowledge. But does the dependence response, or FI, advance the dialectic in any helpful way? One might think that if the dependence response is merely a version of some well-worn response, such as Ockhamism or multiple-pasts compatibilism, then the dependence response has failed to advance the dialectic in any helpful way. For example, Fischer and Tognazzini (2014, pp. 363-365) argue that charitably reconstructed versions of the dependence response must fall into one of these two camps (Ockhamism or multiple-pasts compatibilism) and so there is nothing new in the dependence response. In a similar vein, Patrick Todd and John Martin Fischer (2013, pp. 294-296) argue that an obvious step in the dependence response "crucially involves the notion of dependence" (2013, pp. 295), a notion that both Ockhamists and multiple-pasts compatibilists have sometimes focused on. What, then, does the dependence response have to offer over these more traditional responses?

I am happy to grant that the dependence response must fall into one of these two camps (although I think multiple-pasts compatibilism is a more natural fit), but I'll argue that the dependence response advances the dialectic either way. To show this, let's examine both an Ockhamist version of the dependence response and a multiple-pasts compatibilist version. We'll see that the dependence response presents a distinct and more plausible take on both Ockhamism and multiple-pasts compatibilism.

In their response to the argument from divine foreknowledge, Ockhamists begin by distinguishing between "hard" or "temporally intrinsic" facts about the past and "soft" or "temporally extrinsic" facts about the past. To return to our favorite example, the fact that JFK was assassinated is a hard fact about the past, but the fact that JFK was assassinated 57 years before I wrote this chapter is a soft fact about the past. Ockhamism grants that every hard fact about the past is fixed for (present) agents, but denies that every soft fact is fixed for (present) agents. Ockhamism then claims that God's past beliefs are soft facts about the past and, thus, are not necessarily fixed for (present) agents.

How, exactly, are we to distinguish between hard and soft facts about the past? What criterion should we use? Although I reject this proposal, it is nevertheless open to the proponent of the dependence response to claim that soft facts about the past just are those that depend on what occurs at later times, whereas hard facts about the past do not depend on what occurs at later times.[5]

[5] This is one of the interpretations Fischer and Tognazzini give to the dependence response:

> One way to understand Ockhamism is as the general view that, once we figure out how best to distinguish between JFK's assassination and the fact that JFK was assassinated 49 years before we wrote this paper, we'll be able to block the

This construal of the dependence response would then count as a version of Ockhamism.

Although a version of Ockhamism, this dependence response would be distinct from and superior to traditional versions of Ockhamism. The traditional way to distinguish between hard and soft facts appeals to the "entailment criterion of soft facthood." According to the entailment criterion, soft facts about the past just are those facts that entail some (contingent) fact about the future, roughly.[6] For instance, the fact that JFK was assassinated 57 years prior to my writing this chapter is a soft fact because it entails that I write this chapter after JFK's assassination. In contrast, the fact that JFK was assassinated in 1963 does not entail that any (contingent) fact obtain after 1963 and, hence, is a hard fact.

But there is a major problem for the entailment criterion, namely, that the notion of entailment is too weak. To borrow a case from Patrick Todd (2013), God's past *decrees* concerning what will happen in the future seem to entail contingent facts about the future – if, 1,000 years ago, God *decreed* that you read this chapter today, that seems to entail that you read this chapter today. But, arguably, God's past decrees are not soft facts about the past. At the very least, it seems that God's

> incompatibility argument by pointing out that God's past beliefs are more like the latter than they are like the former. One way to draw the distinction is in terms of *temporal intrinsicality*, but perhaps another is in terms of *dependence*. On this way of articulating the Ockhamist project, Merricks, McCall, and Westphal (on our charitable revision of their arguments) are just articulating a version of Ockhamism. (2014, p. 364)

[6] This is an oversimplification, but it will suffice for our purposes. See the papers in Fischer (1989) for various versions of this criterion, and see Todd (2013) for critical discussion.

past decrees are fixed or beyond our control. Ockhamism, in conjunction with the entailment criterion, does not capture either of these claims.

This problem is obviated by the dependence version of Ockhamism, however, since God's past decrees do not seem to explanatorily depend on what happens at later times. If anything, it goes the other way: what happens at later times seems to explanatorily depend on God's past decrees. Hence, this version of Ockhamism implies that God's past decrees are hard facts and, thus, fixed for us. Not only, then, is the dependence version of Ockhamism distinct from the traditional version, but it also seems superior to it.

Multiple-pasts compatibilists respond to the argument from divine foreknowledge in a very different way, namely, by denying that every hard fact about the past is necessarily fixed for (present) agents. On this view, it is at least possible that we have the freedom to perform actions that would require the hard past to be different, facts about God's past beliefs perhaps being an example. Under this version, the dependence response would claim that facts about God's past beliefs, although hard facts, are not necessarily fixed because such facts are dependent on the agent's current behavior.[7] This response would count as a version of multiple-pasts compatibilism.

[7] This is another interpretation Fischer and Tognazzini give to the dependence response:

> On second thought, someone might think that there is indeed something new here. That is, someone might think that what is new is the contention that when a hard fact about the past depends in a certain way on an agent's present behavior, then the past fact – even a hard fact – is not fixed. This would then suggest a kind of

Although a version of multiple-pasts compatibilism, this construal of the dependence response is distinct from and superior to traditional versions of multiple-pasts compatibilism. The traditional way to argue that the hard past is not necessarily fixed involves appealing to examples with a very specific structure in which an agent (apparently) can do something that, in the circumstances, seems to require a change in the hard past.[8] Consider John Martin Fischer's example of the salty old seadog:

> **Salty Old Seadog:** Each morning at 9:00 a.m. (for the past forty years) he [the salty old seadog] has called the weather service to ascertain the weather at noon. If the "weatherman" says at 9:00 that the weather will be fair at noon, the seadog goes sailing at noon. And if the weatherman says that the weather won't be fair at noon, the seadog *never* goes sailing at noon. The seadog has certain extremely regular patterns of behavior and stable psychological dispositions – he is careful to find out the weather forecast, is not forgetful, confused, or psychologically erratic, and whereas he loves to go

> "defense" of multiple-pasts compatibilism (although only in the context of debates about God's foreknowledge, not causal determinism). (2014, p. 365)

Fischer and Tognazzini go on to argue that this view does not address the Principle of the Fixity of the Past. That might be true of some presentations of the dependence response, but not mine. In fairness to Fischer and Tognazzini, they were engaged with authors who had presented the dependence response in a less than perspicuous way.

[8] For the classic example, see Saunders (1968); see Fischer (1994, ch. 4; 2016, ch. 5) for discussion.

> sailing in sunshine, he detests sailing in bad weather. (1994, pp. 80-81)

Now suppose that the seadog was told at 9:00 this morning that the weather at noon would be horrific, and so he does not in fact go sailing at noon. Nevertheless, we might ask: could the seadog have gone sailing at noon? Plausibly so. However, this is *despite* the fact that the following "backtracker" (a conditional whose consequent concerns the hard past) is true: if the seadog had gone sailing at noon, he would have received a different forecast at 9:00. So, this seems to be a case where an agent is free to do otherwise despite the fact that doing so seems to require (in some sense) that the hard past be different.

There are several problems with this case, though. First, as Fischer (1994, pp. 81-85) argues, it is controversial both whether the agent really is free to do otherwise in such examples and whether the relevant "backtracker" is true. If someone denies that the seadog could have gone sailing, or that the "backtracker" is true, what more could the multiple-pasts compatibilist say? To use a helpful phrase from Fischer, using such examples seems to lead to a "dialectical stalemate."

This problem is obviated by the dependence version of multiple-pasts compatibilism. For one, the dependence response relies on intuitive claims about the order of dependence rather than on controversial claims about such examples.[9]

[9] One might worry that the dependence version gives rise to a new problem – and thus is not superior to the traditional version – since it presupposes that it is possible for the hard past to

For example, consider the case for FI presented in the first chapter: few of the arguments there rely on terribly controversial cases like the salty old seadog (I think). On the contrary, many of the arguments for FI presented there *assumed* that the hard past is (generally) fixed for present agents. For this reason, I think it is fairly clear that the case for the dependence version of multiple-pasts compatibilism is stronger, or at least dialectically richer, than the standard arguments for more traditional versions.

There is another, deeper worry for the **Salty Old Seadog** case as well. Even if successful, it only casts doubt on a particular formulation of FP. Specifically, it casts doubt on the following formulation (called "FP-CF" for "Fixity of the Past-Counterfactual"):

> FP-CF: An agent *S* can do otherwise than *X* at *t* only if, had *S* done otherwise than *X* at *t*, the intrinsic (or hard) past relative to *t would not* have been different.

But, as John Martin Fischer (2011) notes, this is not the only conception of the fixity of the past. Consider the following conception (called "FP-PW" for "Fixity of the Past-Possible World"):

> FP-PW: An agent *S* can do *X* at *t* only if there is a possible world with the same intrinsic (or hard) past up to *t* in which *S* does *X* at *t*.

depend on what occurs later on, and this is controversial. But even if this is right, discussion of this worry would take the dialectic in a new direction, allowing for dialectical progress, and would not, like the traditional version of multiple-pasts compatibilism, lead to a dialectical stalemate.

While the **Salty Old Seadog** case may impugn FP-CF, it does not impugn FP-PW. Even if the *closest* possible world where the seadog goes sailing at noon is a world with a different hard past, it seems implausible to suppose that *every* world where the seadog goes sailing at noon is a world with a different hard past. But that latter claim is exactly what we need if the **Salty Old Seadog** case is to be a counterexample to FP-PW.

Now here's the crucial point: the argument from divine foreknowledge looks just as plausible using FP-PW instead of FP-CF. It does seem that *every* world where *S* does otherwise than *X* is a world where God has a different past belief. Hence, if the argument is formulated with an appeal to FP-PW instead of FP-CF, the **Salty Old Seadog** case fails to undermine the argument. In contrast, by claiming that no dependent fact, temporally intrinsic or otherwise, is necessarily fixed for the present agent, the dependence response undermines FP-PW as well as FP-CF. By doing so, the dependence response seems to be considerably stronger than traditional versions of multiple-pasts compatibilism.

Maybe this is too quick though. Perhaps the traditional multiple-pasts compatibilist can offer an alternative case that does seem to be a counterexample to FP-PW. Consider a classic case from Alvin Plantinga:

> **Paul and the Ant Colony:** Let's suppose that a colony of carpenter ants moved into Paul's yard last Saturday. Since this colony hasn't yet had a chance to get properly established, its new home is still a bit fragile. In particular, if the ants were to remain and Paul were to

> mow his lawn this afternoon, the colony would be destroyed. Although nothing remarkable about these ants is visible to the naked eye, God, for reasons of his own, intends that it be preserved. Now as a matter of fact, Paul will not mow his lawn this afternoon. God, who is essentially omniscient, knew in advance, of course, that Paul will not mow his lawn this afternoon; but if he had foreknown instead that Paul *would* mow this afternoon, then he would have prevented the ants from moving in. (1986, p. 254)

Plantinga claims that, intuitively, Paul can mow his lawn this afternoon despite the fact that his doing so would require the truth of a certain backtracker. But, arguably, it requires even more. If we think of God's past beliefs and intentions as temporally intrinsic (or hard) facts, as seems plausible, this also seems to be a case where Paul can perform an action despite the fact that there is *no* world with the same intrinsic past where he performs that action: arguably, *every* world where Paul mows his lawn is either a world where God doesn't intend on preserving the ants or a world where the ants never moved into Paul's yard in the first place.[10] Hence, this seems to be a counterexample to FP-PW. Perhaps traditional multiple-pasts compatibilists can motivate a rejection of FP-PW just as well as the dependence response theorist can.

[10] To be clear, Plantinga does not think of God's past intentions or beliefs as temporally intrinsic or hard facts.

Even if **Paul and the Ant Colony** successfully motivates a rejection of FP-PW, there are still two respects in which the dependence response has the advantage over traditional multiple-pasts compatibilism. First, the dependence response can *explain* why we might think that Paul is free to mow his lawn.[11] Arguably, the fact that the ant colony is currently in his yard is partly explained by Paul's decision to not mow his lawn this afternoon. After all, the ants moved into his yard partly because God believed Paul would not mow his lawn, and God believed that because Paul decides not to mow his lawn. Hence, according to the dependence response, neither the presence of the ant colony nor God's relevant past mental states are necessarily fixed for Paul.[12] By offering an explanation, the dependence response advances the dialectic beyond traditional multiple-pasts compatibilists who merely appeal to intuition about such cases.[13]

The second point to note is that the dependence response can treat the **Salty Old Seadog** case differently than **Paul and the Ant Colony**. Given that the weather forecast is not explained by what the seadog does at noon, the dependence response implies that the weather forecast is fixed for him at noon. But given that the presence of the ant colony is explained by Paul's refraining from

[11] In places, it seems as if Plantinga tries to offer an explanation. However, his explanation seems to be in terms of entailment which, as I argued above, is problematic.

[12] See Swenson (2017) for further discussion of FI and its relation to this case. There may be a way of construing the case where neither the presence of the ant colony nor God's past mental states depend on Paul's decision to not mow his lawn. In such a case, though, it is far less obvious that Paul is free to mow his lawn. See the introduction in Fischer & Todd (2015) for a nice overview of the issues around the notion of dependence, as well as Wasserman (forthcoming).

[13] See Todd & Fischer (2013) for further discussion about this case and its relation to various versions of the Principle of the Fixity of the Past.

mowing his lawn, the dependence response insists that the presence of the ant colony is not necessarily fixed. Treating these two cases differently is a novel implication of the dependence response, as these two cases (or similar ones) are often lumped together by traditional versions of multiple-pasts compatibilism. For both of these reasons, I conclude that the dependence version of multiple-pasts compatibilism is distinct from and superior to traditional versions of the view.

Summing Up

I have argued that the dependence response does not suffer any dialectical impropriety. Or, at the very least, the dependence response as I have presented it, does not. I would like to make a concession, though. The dependence response goes beyond a mere endorsement of something like FI insofar as it also accepts the claim that God's past beliefs are dependent, in the relevant way, on the agent's present behavior. While this claim may seem fairly intuitive, it is a bit mysterious how this could be. After all, we don't think of *our* past beliefs as depending on any agent's future behavior. How, then, could God's?[14]

Sometimes authors will offer an analogy here. Just as we have access to different parts of *space*, so God has access to different parts of *time*: while telescopes allow us to observe distant places, perhaps God has a "time telescope"

[14] Again, see Fischer & Todd (2015) as well as Wasserman (forthcoming) on the issue of how the dependence theorist ought to think of God's past beliefs as depending on the future.

which allows him to observe distant times. While this is an evocative analogy, it really doesn't help much. How could God have access to different times? What mechanism gives God such access? Consider how ordinary telescopes work: they use mirrors and lenses to amplify light received from distant places. How in the world would this work in the case of a "time telescope"? The analogy seems to break down considerably.[15]

I must admit that I do not have well-developed answers to these questions. It is simply not obvious *how* God's past beliefs could be dependent, in the right way, on any agent's present behavior. But I do think it is at least somewhat plausible *that* this is so (supposing God exists). Nevertheless, to settle whether divine foreknowledge undermines our freedom, more work needs to be done. In particular, we need to understand by what mechanism God's past beliefs could depend, in the relevant way, on an agent's present behavior. Unfortunately, that will have to wait for later work.[16]

However, I would like to make a small point. It seems to me that the unclarity surrounding the mechanism involved in God's foreknowledge can explain some of our other intuitions. Consider the argument from future

[15] Thank you to Michael Nelson for pressing this point.

[16] The issue of the "mechanism" of God's foreknowledge is a long-standing issue. See Boethius (523/1969), Molina (1588/1988) , and Edwards (1845/2009) for classical discussions; see Alston (1986), Craig (1987), the introduction in Fischer (2015), and Wasserman (forthcoming) for contemporary discussions. As a *very* gestural remark, it seems to me that one promising, albeit underexplored, strategy involves thinking of God as occupying more than our four-dimensional spacetime – as occupying a "hyperspace" or even a "hypertime" *in addition* to the more familiar dimensions of space and time. Under this suggestion, God wouldn't be "outside" of space and time, but rather "beyond" space and time. See Hudson (2005; 2014) for discussion of related issues.

contingents again. While very few authors find this argument compelling, considerably more authors find the argument from divine foreknowledge so. And there seems to be something to this discrepancy: it has always been somewhat intuitive, at least to me, that divine foreknowledge would be a more serious threat to our freedom than future contingents. FI gives a nice explanation here. It seems extremely plausible, perhaps even obvious, that certain future contingents are (or would be) at least partly explained by the agent's future behavior. And given an appropriate model of time, we can provide a mechanism as to how this works: given the popular, albeit somewhat controversial view that all times are equally real, it is relatively straightforward how certain past facts can be explained by future facts. By comparison, it may be *somewhat* plausible that certain of God's past beliefs are (or would be) partly explained by the agent's future behavior. But it doesn't seem as plausible as in the case of future contingents. At the very least, providing a mechanism as to how this works is not nearly as straightforward. Assuming something like FI is correct, this difference can explain why we might find the threat from divine foreknowledge more troublesome than the threat from future contingents.

FI and Causal Determinism

In contrast to future contingents and divine foreknowledge, there are many kinds of arguments for the claim that the freedom to do otherwise is incompatible with (causal) determinism. But perhaps the most influential kind are so-called "consequence style" arguments.[17] Here's one way that such arguments go. Suppose that *S* performs action *X* at time *t* and that determinism is true. If determinism is true, then the intrinsic past (relative to *t*), in conjunction with the laws of nature, entails that *S* performs action *X* at *t*. So, if *S* could have done otherwise than *X* at *t*, then either *S* could have performed an action that would have required the past to be different or *S* could have performed an action that would have required the laws to be different. But no agent can perform an action that requires the past to be different nor can an agent perform an action that requires the laws to be different. So, if *S* performs action *X* at *t* and determinism is true, then *S* couldn't have done otherwise than *X* at *t* – freedom and determinism are incompatible. Following our convention, let's call this the "argument from determinism."[18]

Once again, the argument in question invokes FP, a principle which, if I'm right, ought to be rejected in favor of FI. But there's an important difference between the argument from determinism and the previous two arguments, namely, if we rerun the argument with the appeal to FP being replaced by an appeal to FI, the argument *still seems plausible*. Here's how the modified argument would go: suppose that *S* performs action *X* at *t* and that determinism is true. If *S* could have done otherwise than *X* at *t*, then either *S* could have performed an action that would have required the past to be different or *S* could have performed an action

[17] See van Inwagen (1983), Ginet (1990), and Fischer (1994).

[18] The other important families of arguments include "manipulation" arguments and "ultimacy" arguments. See Mele (2006) and Strawson (1994), respectively, although such authors are typically more concerned with the relation between moral responsibility and causal determinism.

that would have required the laws to be different. Now, neither the (intrinsic) past nor the laws are explained whatsoever by what *S* does at *t* – both the intrinsic past and the laws are *explanatorily independent* of *S*'s behavior at *t*. So, if *S* could have done otherwise than *X* at *t*, then *S* could have performed an action that would have required something which is *explanatorily independent* of her behavior to be different. But no agent can perform an action that requires something explanatorily independent of her behavior to be different (i.e. FI is true). So, if *S* performs action *X* at *t* and determinism is true, then *S* couldn't have done otherwise than *X* at *t* – freedom and determinism are incompatible.

In addition to invoking FI, the important difference between this modified version of the argument and the original is the following claim: that neither the intrinsic past nor the laws are explained whatsoever by what *S* does at *t*. At least in our own case, this seems eminently plausible: my writing this chapter right now, say, in no way explains the state of the universe millions of years ago nor does it explain why the laws of nature are what they are. But let's spend a little more time here.

Let's start with the claim that the laws of nature are in no way explained by our current behavior. This seems to presuppose a so-called "governing" conception of lawhood: that the laws genuinely "govern" the way the world unfolds. The alternative is a so-called "Humean" conception of lawhood according to which the laws don't genuinely "govern" the way the world unfolds; the laws are instead generalizations (of a certain type) that merely *describe* how the world unfolds. For instance, consider the law that nothing travels faster than the speed of light. According to the governing conception of lawhood, none of us are traveling faster than the speed of light *because* of this law. In contrast, the Humean conception seems to imply that this law holds partly *because* none of us are traveling faster than the speed of light. After all, if the law that nothing travels faster than the speed of light is a generalization that merely *describes* how fast things move, then this law cannot genuinely *explain* why any object in particular does not travel faster than the speed of light.[19]

So, it seems that if the Humean conception of lawhood is correct, even the modified argument for the incompatibility of freedom and determinism fails. In my view, this is exactly the right result. As others have pointed out, if Humeanism is right, it seems implausible that the laws are fixed or beyond our control.[20] And, once again, FI goes one step further in providing an *explanation* for why this is so. FI insists that it is not the laws per se that are fixed or beyond our control; rather, it is any fact which is beyond our explanatory reach. Insofar as the debate over Humeanism is partly a debate over whether the laws are within our explanatory reach, FI makes it clear why the debate is relevant to freedom.

[19] For an anti-Humean view of the laws, one that I am especially sympathetic with, see Maudlin (2007, chs. 1 and 2); for an influential Humean view of the laws, see Lewis (1986; 1994).
[20] See Beebee & Mele (2002).

Now, I reject the Humean conception of lawhood for reasons we need not consider here.[21] I merely wish to note how FI sheds new light on an issue that has been around for quite some time.

Let's move on to the claim that the intrinsic past, say the intrinsic state of the universe a million years ago, is not explained by our current behavior. While this claim does seem obviously true, it also only seems to be *contingently* true. If we had access to a time machine, or if the direction of causation ran from future to past, the past would not necessarily be beyond our explanatory reach. Hence, the argument from determinism, once modified as to include FI, only establishes that determinism is incompatible with freedom *for agents in situations similar to ours*.

This theme has been picked up by several authors recently. Carolina Sartorio puts the point nicely:

> Our discussion of the contingency objection to the consequence argument suggests that the so-called problem of determinism and free will is not the problem of determinism and free will. The threat to our free will is not just determinism; it is determinism *plus some additional fact*: a contingent fact about us. (2015, pp. 258-259)

This seems exactly right to me. The thesis of determinism isn't threatening *merely* because, if true, there is some fact, or set of facts – in this case, a description of a state of the universe and the laws – that entails our behavior. After all, if there are

[21] Again, the most forceful considerations, in my view, come from Maudlin (2007, chs. 1 and 2).

future contingents, or if there is a god with exhaustive divine foreknowledge, then there are also facts that entail our behavior. Or as Sartorio (2015, p. 258) notes, if determinism is true, then there is also a *future* state of the universe that, in conjunction with the laws, entails our present behavior: given, say, the intrinsic state of the world one million years from now, and the laws, there is only one way that the intrinsic state of the world could be right now. But we don't think *this* entailment relation is a threat to our freedom. Rather, the thesis of determinism is threatening because, if true, there are facts which are *completely beyond our control* – facts about the *past* and the *governing* laws – that entail our behavior. These facts may not be beyond every *possible* agent's control, but they seem to be beyond ours.

So, FI only establishes a fairly modest conclusion: that the freedom to do otherwise is incompatible with determinism *for agents in a situation similar to ours*. Still, there are many who reject even this minimal conclusion, namely, so-called "classical compatibilists." Classical compatibilists hold that the freedom to do otherwise is compatible with determinism, even for agents in our situation. Should these authors feel any pressure to endorse FI and its implications?

The issue is a complicated one, and the complications are further explored in the appendix of this dissertation. But in short the answer is "It depends." For instance, perhaps the most influential version of classical compatibilism, so-called "local-miracle compatibilism," seems to accept the principle of the Fixity of the

Past (FP) or something close to it.[22] It is the principle of the Fixity of the Laws (FL) which they reject. But many of the arguments I gave for FI start with the plausibility of FP. Insofar as these arguments are compelling, it would seem that those who accept local-miracle compatibilism should feel some pressure to endorse FI. In contrast, another version of classical compatibilism, so-called "multiple-pasts compatibilism," straightforwardly rejects FP.[23] (Importantly, this understanding of "multiple-pasts compatibilism" accepts the compatibility of freedom and physical determinism, whereas the understanding of "multiple-pasts compatibilism" presented in the last section accepts the compatibility of freedom and divine foreknowledge; both understandings reject a *strict* interpretation of FP, though.) Without much sympathy toward FP whatsoever, the arguments I presented for FI do not even get off the ground. Thus, whether the classical compatibilist should feel any pressure to accept FI depends on what version of classical compatibilism she endorses.

I'm willing to accept this result. But in closing this section, I would like to sketch a more general argument against classical compatibilism, one that promises to reshape the dialectic in a helpful way.

Let me start with a way of framing debates over the freedom to do otherwise, one that I think is quite common, albeit not explicitly formulated. So far, I have

[22] See Lewis (1981) for the classic presentation. See Vihvelin (2013) for a more modern take. The term "local-miracle compatibilism" is from Fischer (1988; 1994)
[23] See Saunders (1968) for the classic presentation. See Dorr (2016) for a more modern rejection of FP (or something closely related).

been using the notions of a fact being “beyond the agent’s control” and a fact being “fixed for the agent” interchangeably, as is often done. But it will prove helpful to think about the relationship between these two notions. The former is straightforwardly an agential notion, being concerned with a certain kind of control. But the notion of a fact being “fixed for the agent” might be a bit broader, as it uses the notion of “fixity” which is closely connected to possible world semantics. Here are two examples where the notion of “fixity” plays a pivotal role.

In the widely popular Lewis-Stalnaker semantics for counterfactuals, we are told that a subjunctive conditional of the form “If *P* were true, then *Q* would be true” holds just in case the closest world where *P* is true is a world where *Q* is also true (roughly).[24] What makes one world “closer” than another? That is a tricky question,[25] but there is a large (if not unanimous) consensus that context often affects the evaluation. For the first example, suppose that Alex weighs 150 pounds and Austin weighs 200. Now consider these two counterfactuals:

1. If Alex weighed as much as Austin, their combined weight would be in excess of 300 pounds.
2. If Austin weighed as much as Alex, their combined weight would not be in excess of 300 pounds.[26]

[24] See Stalnaker (1968) and Lewis (1973), although there are important differences.
[25] See the appendix for discussion.
[26] These examples have been slightly modified from Maudlin (2007, ch. 1).

While both of these counterfactuals seem to be true, it's initially puzzling how this could be. At first glance, it would appear that "Alex weighs as much as Austin" expresses the same proposition as "Austin weighs as much as Alex," since the "same weight" relation is symmetric. Hence, it would appear that the closest world where one is true is the closest world where the other is true. But then how can both counterfactuals be true?

The answer is that, given standard English conventions, the claims "Alex weighs as much as Austin" and "Austin weighs as much as Alex" pick out different propositions. As English speakers, we've decided that order matters in this case. When we are considering the claim "Alex weighs as much as Austin," we know to keep Austin's actual weight the same, and to vary Alex's weight to match; and just the opposite when we are considering the claim that "Austin weighs as much as Alex." To use the terminology, context tells us whose weight to hold *fixed* in evaluating the counterfactual scenario.

The second example is closer to home. Suppose a famous pianist, Meghan Watson, is sitting in front of a piano but with her hands tied down. Can she play the piano? We are somewhat ambivalent. On the one hand, of course she can – she's Meghan Watson! But on the other hand, of course she can't – her hands are tied down! In her classic paper, Angelika Kratzer (1977) provides a systematic account of what's happening here, one that again appeals to possible worlds and the notion of "fixity." Kratzer suggests that claims of the form "Agent *S* can perform action *X*" are equivalent to the claim "Holding fixed the relevant facts, it is possible that *S* perform action *X,*" where what counts as a "relevant fact" is determined by context. When we are inclined to say that Meghan can play the piano, we are in a context where we don't deem her hands being tied down as relevant and, hence, don't hold that fact fixed. And if we don't hold that fact fixed, then it surely seems possible that she play the piano. However, when we are inclined to say that Meghan can't play the piano, we are in a different context, one where we do deem her hands being tied down as relevant. And if we do hold that fact fixed, it doesn't seem possible for her to play the piano. Again, context tells us what facts to hold *fixed,* this time when evaluating claims about what an agent can do.

So, the notion of a fact's being "fixed" is fairly broad notion, one that is useful in many different areas. But what about the notion of a fact's being fixed *for an agent*? What are we to make of this more substantial notion? I'd like to suggest that, in some ways, *this is the central question of freedom*. Here's what I mean. While what an agent "can" do and what the agent is "free" to do are importantly related, they aren't synonymous. Think about Meghan Watson's case again. While we may be ambivalent as to whether she *can* play the piano, we are in no way ambivalent about whether she is *free* to play the piano. At least in my own case, there is simply no pull toward saying she is *free* to do so. Does that mean "can" and "freedom" are completely unrelated? Not necessarily. Instead, I prefer to think of "freedom" as a particular *restriction* of "can" – that when we are determining whether an agent is *free* to perform a certain action, we are asking whether she can perform the action *given the context of freedom*. That is, by speaking of freedom, we are automatically given a certain context where we have some idea as to what facts are to be held fixed and which ones aren't. Think about the case of Alex and Austin again. When we consider the claim that "Alex weighs as much as Austin," we know to hold a certain fact fixed, namely Austin's actual weight, and let others vary. Likewise, when we consider the claim that "Meghan is free to play the piano," we know to hold certain facts fixed, like the fact that her hands are tied down, and perhaps let others vary. The central question about freedom then becomes this: *in the context of freedom*, *which facts are fixed*?

While few authors are this explicit when setting up various debates about freedom, something like this framework is often in the background. For instance, many authors claim that in the various debates over freedom, we are interested in a particular sense of "can," one that is relevant to moral responsibility, or promise-making, or practical deliberation, etc. The framework I am offering is a natural fit here. Under my suggestion, these authors hold that talk about freedom, which may also be picked out by talk about moral responsibility, or promise-making, or practical deliberation, etc., isolates a certain context for evaluating a "can" claim. The debate is then over what facts are fixed in that context. Incompatibilists about the freedom to do otherwise and determinism hold that facts about the past as well as facts about the laws are fixed; local-miracle compatibilists hold that some facts about the laws aren't fixed; multiple-pasts compatibilists hold that some facts about the (intrinsic) past aren't fixed.

Here's how this framework helps. By framing the debate this way, we can see that, despite important differences, local-miracle compatibilism and multiple-pasts compatibilism share a deep similarity: both views claim that just because a fact is *beyond our control*, it does not follow that the fact is *fixed* in determining what we are *free* to do. Let's start with the case of local-miracle compatibilism. Typically, these authors are willing to grant that the laws are beyond our control in some important sense. For instance, David Lewis writes:

> …consider what really would be a marvelous ability to break a law – an ability I could not credibly claim. Suppose that I were able to throw a stone very, very hard. And suppose that if I did, the stone would fly faster than light, an event contrary to law. Then I really would be able to break a law. For starters: I would be able to do something such that, if I did it, a law would be broken. But there is more to be said. I would be able to do something such that, if I did it, my act would *cause* a law-breaking event. (1981, p.115; emphasis added.)

In this passage, Lewis seems to admit that the laws are not "up to us" in any deep way. If they were, then we would have the ability to perform an act that would either *cause* or *in itself be* a law-breaking event. We would then have a choice, in a very deep sense, as to which laws actually obtain.[27] But, as Lewis puts it, he cannot

[27] One might grant that, in this causal sense, Lewis isn't committed to saying that the laws are "up to us," but that they are still "up to us" in a problematic sense. For instance, consider a counterfactual account of "up to us": a fact, *F*, is "up to" agent *S* if, and only if, there is some

credibly claim that we have such an ability. Instead, Lewis says he is merely committed to the following claim:

> Had I raised my hand, a law would have been broken before-hand. The course of events would have diverged from the actual course of events a little while before I raised my hand, and at the point of divergence there would have been a law-breaking event – a divergence miracle, as I have called it. … To accommodate my hypothetical raising of my hand while holding fixed all that *can and should be held fixed,* it is necessary to suppose one divergence miracle, gratuitous to suppose any further law-breaking. (pp. 116-117; emphasis added.)

Since Lewis is committed to the claim that he is free to raise his hand, he also seems committed to the claim that the laws *need not be held fixed* in evaluating what we are free to do. The (distant) past, yes, but not the laws. And when taken with the previous passage, we get an interesting result: Lewis seems committed to the claim that the laws are completely beyond our control and, yet, Lewis also seems committed to the claim that the laws need not be held fixed in evaluating what we are free to do.

action, *X*, that *S* is free to perform such that, if *S* were to perform *X*, *F* would not have obtained. If this account is correct, then Lewis's view implies that the laws are "up to us." But this account is controversial. For the sake of argument, I am willing to grant Lewis and other classical compatibilists their favored account of a fact being "up to us." Thank you to John Fischer for this point.

Multiple-pasts compatibilists say something similar about the past. John Turk Saunders puts the point like this:

> Someone may confusedly think that since the power to perform an act that is empirically sufficient for a future situation is the power to *cause* that situation, then the power to perform an act that is empirically sufficient for a past situation must likewise be the power to *cause* that past situation. This confusion may lead one to think that the power so to act that (to perform an act such that if performed) a past situation would have taken place is a power to *cause (or to change)* the past. But it is not. It is only the power to do something that one would do *only if* a certain situation had taken place in the past. (1968, p. 102; emphasis added.)

According to Saunders, we cannot *cause* or *change* the past and, thus, it would seem that the past is not "up to us" in any deep way. Nevertheless, we sometimes have the power to act in such a way that *requires* a difference in the past. Thus, just like Lewis, Saunders also seems committed to the claim that there are facts, facts about the past in this case, that are beyond our control, but that these facts need not be held fixed in evaluating what we are free to do.

So, views like Lewis's and Saunders's have a deep similarity, namely, they both deny the following principle: if fact *F* is completely beyond *S*'s control at time *t*, then *F* is fixed in evaluating what *S* is free to do at *t*. But once we see this similarity, we might start to wonder about the plausibility of such views. The inference from lack of control to fixity is quite natural. To see this, notice that something analogous seems to hold in much of our practical reasoning. Consider an example from John Martin Fischer cited in the previous chapter:

> **Icy Patch**: Sam saw a boy slip and fall on an icy patch on Sam's sidewalk on Monday. The boy was seriously injured, and this disturbed Sam deeply. On Tuesday, Sam must decide whether to go ice-skating. Suppose that Sam's character is such that if he were to decide to go ice-skating at noon on Tuesday, then the boy would not have slipped and hurt himself on Monday. (Fischer 1994, p.95)

Now imagine that Sam is trying to decide whether or not to go ice-skating. The following reasoning seems deeply suspect: "Well, if I were to go ice-skating, the accident wouldn't have occurred. That would be great for the boy who slipped. So, I should go ice-skating." What explains why this line of thinking is so problematic?

In the previous chapter, we examined whether FP or FI was a better explanation. But we need not take a stand here. Rather, there seems to be a more general explanation, namely, that the boy's injury is completely beyond Sam's control and, hence, ought to be held fixed in Sam's deliberation. FP and FI are rival accounts as to *why* the boy's injury is completely beyond Sam's control, but that's all. Whether FP or FI is invoked to explain the speciousness of Sam's reasoning, both explanations rely on the inference from lack of control to fixity. Allow me another example:

> **Victor's Vacation:** Victor has some vacation time this week and very badly wants to spend it in one of his favorite places in the world, Joshua Tree National Park. The only feasible way he has to get there is to drive his car which would take a little over an hour. Victor is a very careful and thorough individual and always makes sure his car is in working order before driving it. Unfortunately, Victor discovers his car is currently in terrible shape, so terrible that if he were to drive it for more than 30 minutes in its current condition, it would explode.

Now compare two variants on this case. First, Victor has absolutely no way of fixing his car. He's taken it to several reliable mechanics, all of whom have told him that the car is simply beyond repair. That is, the current state of his car is beyond his control. What should Victor decide to do in this case? Obviously, it's not the case that he ought to decide to travel to Joshua Tree this weekend – in fact, his traveling to Joshua tree is not even *an option* in this scenario. Holding fixed that he can only get there using his car, and that his car is beyond repair, the possibility of traveling to Joshua Tree is definitively ruled out.

But now consider the second variant. Everything is exactly the same except that there is a simple fix for his car's condition: all he has to do is refill the fluids and his car will easily make the trip to and from Joshua Tree. What should Victor decide in this case? Presumably, he should decide to make the trip – at the very least, his traveling to Joshua Tree is *an option*.[28]

Why is traveling to Joshua Tree a deliberative option for Victor in the second scenario but not the first? The following reasoning seems quite attractive: only in the first scenario is the unfortunate state of Victor's car *beyond his control*. Hence, in the first case, but only the first case, he ought to hold the state of his car *fixed* in evaluating what to do this weekend. And holding fixed the state of his car, it is not possible for Victor to visit Joshua Tree this weekend.

This line of reasoning assumes the following: if fact *F* is beyond *S*'s control at time *t*, then *F* is fixed in assessing what *S* ought (practically) to do at time *t*.[29] Granted, this is a principle about what one (practically) *ought* to do, but it's tempting to think that the same holds for what one is *free* to do as well.[30] That is, it is tempting to think that if fact *F* is completely beyond *S*'s control at time *t*, then *F* is fixed in evaluating what *S* is free to do at time *t*. Insofar as views like Lewis's and Saunders's deny this line of thinking, their view requires substantial motivation.

[28] What if Victor is unaware of this simple fix? We might say that "subjectively" it's not the case that he ought to go to Joshua Tree, but that "objectively" he ought to. I think a more natural way to put it is like this: his going to Joshua Tree is an option for him, and so he should (in some sense) consider it, but he isn't aware that it's an option.
[29] In the "objective" sense of "ought." See the previous footnote.
[30] Importantly, I do not mean to assume that "ought" implies "can." Rather, I am only suggesting an analogy between practical reasoning and reasoning about freedom. Or at the very least, I am only assuming that the *practical* "ought," not necessarily the *moral* "ought," implies "can."

There are few points worth clarifying here. First, I am not claiming that *all* versions of classical compatibilism face this objection. For instance, so-called "Humean" compatibilists – classical compatibilists who believe that a Humean account of the laws of nature is correct – can accept the inference from lack of control to fixity. Humean compatibilists simply deny that the laws are beyond our control. This objection is only aimed at classical compatibilists who accept that both the laws and the past are beyond our control but that, nevertheless, we are sometimes free to do otherwise.

Second, FI is meant to give a sufficient condition for when a fact is beyond an agent's control, but one need not accept FI to appreciate the objection. One need only be sympathetic to the inference from lack of control to fixity. I believe that FI can provide a unified and deep account of why both the past and the laws are beyond our control and, hence, why both are fixed. But one could deny FI and instead embrace FP, say, while still accepting this argument against these versions of classical compatibilism.

Finally, and most substantively, I do not believe this objection constitutes anything like a decisive objection to classical compatibilism.[31] Rather, I only mean for it to help clarify and move forward the dialectic. Traditionally, the incompatibilist starts by arguing that classical compatibilism implies an

[31] There are all sorts of responses the classical compatibilist could give that are worth examining, but doing so would make this chapter inordinately long. In my view, though, the most plausible response for the classical compatibilist is to adopt some version of "causal decision theory" to explain cases like **Icy Patch** and **Victor's Vacation**. See Weirich (2016) for an overview.

extraordinary claim: either it implies that we are free to break the laws or it implies that we have power over the past (if determinism is true). The classical compatibilist then responds that these claims, when understood correctly, aren't obviously so extraordinary: classical compatibilism only implies the weaker claim that we are free to perform some action such that, if we were to perform it, either the laws or the past would have been different. Incompatibilists then try to show that this weaker claim is still incredible because it violates some basic intuition about the laws or the past. And typically, this is where the dialectic stalls.

I'm suggesting that the dialectic start in a very different way. The incompatibilist should start by pointing out that classical compatibilism seems to imply a more general claim: it implies that there are some facts which, although completely beyond our control, are not (necessarily) fixed in determining what we are free to do. The incompatibilist can then use examples like **Icy Patch** or **Victor's Vacation** to argue that this implication seems dubious. A minimally successful response from the classical compatibilist would involve giving plausible cases where some fact is beyond the agent's control and, yet, is not necessarily fixed for the agent,[32] or even just showing that cases like **Icy Patch** and **Victor's Vacation** do not support the inference from lack of control to fixity. But in order for the classical compatibilist to really be at ease, she would need to (i) give some independently motivated sufficient condition for when a fact beyond the agent's

[32] Various cases involving "back-trackers," like the **Salty Old Seadog** or **Paul and the Ant Colony**, are the only cases I know of that come close to this. Unfortunately for the classical compatibilist, these cases are very controversial – surely more plausible cases are required.

control is not necessarily fixed, and (ii) show that, plausibly, either the laws or the distant past satisfies this sufficient condition. If the classical compatibilist can accomplish (i) and (ii), I believe the incompatibilist should admit defeat.

There are several advantages to framing the dialectic this way. First, it promises to move us away from the debate over what counts as an acceptable formulation of FL or FP. Given how complex and contentious that debate has become, this is highly desirable. Second, it sets clear goals for the classical compatibilist. As mentioned, the classical compatibilist needs to accomplish (i) and (ii) to declare victory over the incompatibilist. Third, and perhaps most importantly, framing the dialectic this way helps us see the deeper issue that classical compatibilists and incompatibilists disagree over. As I have been trying to show, (non-Humean) classical compatibilists are united in rejecting the inference from lack of control to fixity. By isolating this as the source of disagreement, we can hope to make progress in a debate that is infamous for dialectical impasses.

Concluding Remarks

I've argued that, under FI, future contingents and divine foreknowledge are no threat to freedom, but that physical determinism is. The basic difference is an explanatory one: only in the case of determinism do there seem to be facts that are in no way explained by our present behavior which also entail our present behavior. Now, strictly speaking, FI does not *reconcile* freedom with either future contingents or divine foreknowledge. It only undermines the traditional arguments for thinking that freedom is incompatible with these phenomena. And of course, there may be arguments which FI does not undermine. Moreover, there may be other threats to freedom that FI does not engage with. I leave both of these possibilities for future work. What I have said here is contentious enough.

References:

Alston, William (1986). "Does God Have Beliefs?" *Religious Studies* 22: 287-306.

Beebee, Helen & Mele, Alfred (2002). "Humean Compatibilism," *Mind* 111: 201-223.

Boethius (523/1969). *The Consolation of Philosophy*. Translated by Victor E. Watts. London: Penguin Books.

Craig, William L. (1987). *The Only Wise God*. Grand Rapids: Baker Book House.

Cyr, Taylor W. & Law, Andrew (forthcoming). "Freedom, Foreknowledge, and Dependence: A Dialectical Intervention," *American Philosophical Quarterly*.

Dorr, Cian (2016). "Against Counterfactual Miracles," *Philosophical Review* 125: 241-286.

Edwards, Jonathan (1845/2009). *The Freedom of the Will*. Grand Rapids: Soli Deo Gloria.

Finch, Alicia & Rea, Michael (2008). "Presentism and Ockham's Way Out," in *Oxford Studies in Philosophy of Religion*, vol. 1, edited by Jonathan Kvanvig. Oxford: Oxford University Press: 1-17. Reprinted in *Freedom, Fatalism, and Foreknowledge*, edited by John M. Fischer and Patrick Todd. New York: Oxford University Press: 229-246.

Fischer, John M. (1988): "Freedom and Miracles," *Noûs* 22: 235-252.

Fischer, John M. (1989). *God, Foreknowledge, and Freedom*. Stanford: Stanford University Press.

Fischer, John M. (1994). *The Metaphysics of Free Will: An Essay on Control*. Malden: Blackwell Publishing.

Fischer, John M. (2011). "Freedom, Foreknowledge, and the Fixity of the Past," *Philosophia* 39: 461-474.

Fischer, John M. (2015). *Our Fate: Essays on God and Free Will*. Oxford: Oxford University Press.

Fischer, John M. & Tognazzini, Neal A. (2013). "The Logic of Theological Incompatibilism: A Reply to Westphal," *Analysis* 73: 46-48.

Fischer, John M. & Tognazzini, Neal A. (2014). "Omniscience, Freedom, and Dependence," *Philosophy and Phenomenological Research* 88: 346-367.

Fischer, John M. & Todd, Patrick (2015). Introduction to *Freedom, Fatalism, and Foreknowledge*, edited by John M. Fischer and Patrick Todd. New York: Oxford University Press: 1-38.

Ginet, Carl (1990). *On Action*. New York: Cambridge University Press.

Hudson, Hud (2005). *The Metaphysics of Hyperspace*. New York: Oxford University Press.

Hudson, Hud (2014). *The Fall and Hypertime*. Oxford: Oxford University Press.

Kratzer, Angelika (1977). "What 'Can' and 'Must' Can and Must Mean," *Linguistics and Philosophy* 1: 337-355.

Law, Andrew & Tognazzini, Neal A. (2019). "Free Will and Two Local Determinisms," *Erkenntnis* 84: 1011-1023.

Lewis, David (1973). *Counterfactuals*. Cambridge: Harvard University Press.

Lewis, David (1981). "Are We Free to Break the Laws?" *Theoria* 47: 113-121.

Lewis, David (1986). *Philosophical Papers*, vol. 2. Oxford: Oxford University Press.

Lewis, David (1994). "Humean Supervenience Debugged," *Mind* 103: 473-490.

Maudlin, Tim (2007). *The Metaphysics Within Physics*. New York: Oxford University Press.

McCall, Storrs (2011). "The Supervenience of Truth: Freewill and Omniscience," *Analysis* 71: 501–506.

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

Merricks, Trenton (2009). "Truth and Freedom," *Philosophical Review* 118: 29–57.

Merricks, Trenton (2011). "Foreknowledge and Freedom," *Philosophical Review* 120: 567–586.

Molina, Luis de (1588/1988). *On Divine Foreknowledge: Part IV of the "Concordia."* Translated by Alfred J. Freddoso. Ithaca: Cornell University Press.

Origen (246/2002). *Commentary on the Epistle to the Romans: Books 6–10*. Translated by T.P. Scheck. Washington D.C.: The Catholic University of America Press.

Pike, Nelson (1965). "Divine Omniscience and Voluntary Action," *Philosophical Review* 74: 209–216.

Plantinga, Alvin (1986). "On Ockham's Way Out," *Faith and Philosophy* 3: 235-269.

Rea, Michael (2006). "Presentism and Fatalism," *Australasian Journal of Philosophy* 84: 511-524. Reprinted in *Freedom*, *Fatalism*, *and Foreknowledge* (2015), edited by John M. Fischer and Patrick Todd. New York: Oxford University Press: 147-163.

Sartorio, Carolina (2015). "The Problem of Determinism and Free Will Is Not the Problem of Determinism and Free Will," in *Surrounding Free Will*: *Philosophy, Psychology, and Neuroscience*, edited by Alfred Mele. Oxford: Oxford University Press: 255-273.

Saunders, John T. (1968). "The Temptations of 'Powerlessness'," *American Philosophical Quarterly* 5: 100-108.

Stalnaker, Robert C. (1968). "A Theory of Conditionals," in *Studies in Logical Theory*, edited by Nicholas Rescher. Oxford: Basil Blackwell: 98–112.

Strawson, Galen (1994). "The Impossibility of Moral Responsibility," *Philosophical Studies* 75: 5-24.

Swenson, Phillip (2016). "Ability, Foreknowledge, and Explanatory Dependence," *Australasian Journal of Philosophy* 94: 658-671.

Swenson, Phillip (2017). "Fischer on Foreknowledge and Explanatory Dependence," *European Journal for Philosophy of Religion* 9: 51-61.

Todd, Patrick (2013). "Soft Facts and Ontological Dependence," *Philosophical Studies* 164: 829-844.

Todd, Patrick & Fischer, John M. (2013). “The Truth about Foreknowledge,” *Faith and Philosophy* 30: 286-301.

van Inwagen, Peter (1983). *An Essay on Free Will.* New York: Oxford University Press.

Vihvelin, Kadri (2013). *Causes, Laws, and Free Will: Why Determinism Doesn’t Matter*. New York: Oxford University Press.

Wasserman, Ryan (forthcoming). “Freedom, Foreknowledge, and Dependence,” *Noûs*.

Weirich, Paul (2016). "Causal Decision Theory," *The Stanford Encyclopedia of Philosophy* (Winter 2016 edition), edited by Edward N. Zalta. https://plato.stanford.edu/archives/win2016/entries/decision-causal/.

Westphal, Jonathan (2011). “The Compatibility of Divine Foreknowledge and Freewill,” *Analysis* 71: 246–252.

Westphal, Jonathan (2012). “The Logic of the Compatibility of God’s Foreknowledge and Human Freewill,” *Analysis* 72: 746-748.

Chapter 4: Explanation and The Problem of Enhanced Control

Introduction

In the last chapter, I argued that the truth of determinism would undermine our being free to do otherwise. Most days of the week I also happen to think that we are free to do otherwise, at least on some occasions. This combination of claims, let's dub it "libertarianism,"[1] has some puzzling implications (to put it mildly). For one, it seems to suggest that we can know through mere philosophical reflection and a bit of introspection that determinism is false. But determinism appears to be an empirical thesis, something for theoretical physicists to investigate, not philosophers. For another, libertarianism also seems to imply that our conception of ourselves as free, autonomous beings, navigating life's garden of forking paths, is beholden to the "arcane ruminations" of such physicists, to borrow a phrase from John Fischer (2006). But that's at least somewhat bizarre – what does a physicist's mathematical meandering have to do with how we view ourselves in this regard?

The implication of libertarianism that worries me the most though, and the one that is the focus of this chapter, is sometimes called the "problem of enhanced

[1] This definition of "libertarianism" is slightly non-standard, as it is typically defined to imply that the truth of determinism would undermine our *moral responsibility*, something I would like to remain neutral on. Perhaps a better term would be "semi-libertarianism" (to mirror John Fischer's term "semi-compatibilism").

control." It runs roughly as follows. Presumably, if we are free to do otherwise, we have more control over our lives than we otherwise would and, according to libertarianism, our being free to do otherwise requires the falsity of determinism. So it looks as if libertarianism implies that the falsity of determinism – i.e. indeterminism – somehow *enhances* the amount of control we have over our lives. But that's deeply mysterious, if not incoherent. After all, indeterminism merely seems to introduce a kind of *randomness* – that whether we decide one way or another is "up to chance" to some degree. How could adding randomness *increase* or *enhance* the amount of control we have over our lives?

It will be helpful to have an example. To borrow a case from van Inwagen (1983, ch. 4), suppose a thief is sitting in church when the collection plate passes by. The thief is deeply torn over whether to resort to his thievish ways and take the money from the plate or to instead start on a new path and let the money pass by. After a good deal of struggling with himself, his past gets the better of him and he steals from the collection plate before passing it on. If determinism holds in this case, then the state of the world just prior to the thief's decision in conjunction with the laws entails that he decide to steal from the collection plate. According to libertarianism, this means the thief wasn't free to do otherwise and, hence, lacked a certain amount or kind of control in deciding to steal.

But now add a little indeterminism to the mix: suppose instead that the state of the world just prior to the thief's decision in conjunction with the laws doesn't entail that he decide to steal from the plate. Rather, the previous state of the world and the laws only make it more probable than not – they assign an objective probability of 0.7, say, that he decide to steal. Nevertheless, everything happens in exactly the same way: the plate passes in front of him, he struggles immensely, but eventually caves and steals from the plate. According to libertarianism, the addition of indeterminism makes all the difference, granting him the freedom to decide to refrain from stealing and, thus, allowing him to exhibit more control over his decision. But *how*? With indeterminism in the picture, it simply seems somewhat *random* whether he decides to steal or refrain. How does such randomness *increase* his control?[2]

Two points are worth noting. First, it's important to distinguish the problem of enhanced control from the so-called "problem of luck." The former only claims that indeterminism does not *enhance* one's control, whereas the latter goes further and claims that indeterminism *diminishes* it.[3] And by claiming that indeterminism diminishes control, the problem of luck would appear to be a problem for many views of freedom, not just libertarianism. After all, many (if not most) rival views of freedom are willing to accept that neither determinism nor indeterminism necessarily diminishes one's control. But that means such views must face the problem of luck as well. In contrast, it is only the libertarian view which is committed to the claim that indeterminism is *required* for control, which means

[2] This formulation of the problem is most similar to Mele's (2006) formulation, although Mele casts it as a version of the luck objection, not the problem of enhanced control. (See below.)

[3] Hobart (1934) is often taken to be the classic formulation in modern times, although there are a number of distinct formulations. See Franklin (2018, ch. 5) for a helpful overview and critical discussion.

only the libertarian must face the problem of enhanced control. In what follows, I will therefore assume that some sense of control or agency is compatible with indeterminism – that indeterminism does not necessarily undermine the control we have over our decisions, actions, values, etc. Instead, the challenge is to show how indeterminism can *enhance* our control or agency.[4]

Second, the problem of enhanced control is consistent with the claim *that* the freedom to do otherwise is incompatible with determinism. Instead, it merely asks *how* indeterminism is supposed to grant us more control, as libertarianism seems to imply. In other words, the problem of enhanced control challenges the libertarian to come up with a *mechanism* for how indeterminism might increase control.

Here's a way to see this last point. Suppose someone thinks that the freedom to do otherwise is utterly *impossible*. Such an individual will grant that the freedom to do otherwise is incompatible with determinism – since freedom is impossible, it's incompatible with everything! But this individual can still press the problem of enhanced control against the libertarian.[5] It's perfectly sensical for such a person to say something like: "You libertarians admit that determinism is incompatible with freedom, but indeterminism doesn't seem any better. Tell me, exactly, how does a certain kind of randomness gives us *more* control?" So, a satisfactory response to the problem of enhanced control must go beyond arguments that are

[4] Thank you to Kadri Vihvelin and Carolina Sartorio for helping me get clear on this.

[5] I have authors like Strawson (1994) in mind here, although he is concerned with the notion of moral responsibility rather than the freedom to do otherwise.

only meant to show *that* determinism is incompatible with the freedom to do otherwise. Instead, a satisfactory response needs to make it clear *how* indeterminism helps secure our freedom. What is it, exactly, that determinism rules out that indeterminism allows for? And how, exactly, does indeterminism allow for it?

The purpose of this chapter is to provide a preliminary but original answer to this challenge. In short, I'll be arguing that indeterminism (in the right places) allows an agent's decision to play a certain kind of explanatory role in the world, what I'll call an *ineliminable* explanatory role. This explanatory role is a necessary condition on an idea closely related to freedom, if not synonymous with it, namely the idea of *authorship* – of being the *author* of one's life. To put it in a grandiose way, I'll argue that if indeterminism holds (in the right places), then it is impossible to explain the history of the world without citing our decisions, actions, values, etc. If so, then it would seem that indeterminism can add to our agency in an important way.

Before developing my answer to the problem of enhanced control though, I'll first (very) briefly consider other answers that have been given. By reflecting on both the advantages and disadvantages of these previous answers, we'll uncover desiderata for any solution. I'll then explicate my answer and argue that it meets these desiderata in a deeply satisfying way. Finally, I'll conclude by considering an objection unique to my solution.

Previous Solutions to the Problem of Enhanced Control

Agent-Causal Solutions

We typically think of causation as fundamentally involving *events*: that the window broke because Suzy threw a rock at it. But according to some, causation can also fundamentally involve *substances*: that Suzy (full stop) caused the window to break. Importantly, on this view, instances of substance-causation are *not reducible* to instances of causation that merely involve events: Suzy's causing the window to break doesn't "reduce to" or isn't wholly "grounded in" the fact that the window broke because Suzy threw a rock at it or similar facts. And if agents are substances, then a view emerges wherein agents can sometimes play a *fundamental* or *irreducible* role in causing certain events. Several authors have argued that, with perhaps a little tinkering, agent-causation provides the most promising answer to the problem of enhanced control.

Specifically, agent-causal solutions typically endorse two claims: (i) that (free) agents have the power to be fundamentally or irreducibly causally involved in their (free) actions, and (ii) that this kind of power requires indeterminism. Let's return to the thief's decision. If determinism is true, then according to the agent-causal solution, there is no room for the *thief* to play an irreducible causal role in the production of his decision to steal. Instead, it is merely certain mental states – his beliefs, desires, etc. – in conjunction with the environment that bring about his decision. But if his decision is not determined, then his mental states in conjunction with the environment may make his decision *probable*, but not guaranteed. And that extra bit of room is where *he* can make an irreducible causal contribution: *he* can directly bring about the decision to steal from the collection plate in a fundamental and irreducible way.

Of course, this is hardly even a sketch of the agent-causal solution to the problem of enhanced control, as there are many distinct versions of it.[6] But regardless of how it is made more precise, I think we should admit that there is something attractive about it as a possible solution. Specifically, the idea of playing a certain kind of *fundamental* or *irreducible* causal role is quite appealing. We want to be more than the product of our circumstances or experiences; we want to contribute something new and original to the world. By thinking of ourselves as playing a fundamental or irreducible causal role, we seem to be getting at these ideas and, thus, this part of the agent-causal solution looks quite promising.

Despite its promise, though, the agent-causal solution has a number of unappealing features as well. First and foremost is the general notion of substance- or agent-causation. Many authors, myself included, find this notion fairly mysterious. We seem to have a pretty good grip on events causing events; we also seem to have a pretty good grip on agents causing events in a *reducible* or *non-fundamental* way. But it's not clear that we have a grip on what it means for an agent to cause an event in an *irreducible* or *fundamental* way. Indeed, the claim that substance- or agent-causation is *irreducible* or *fundamental* is simply a

[6] For a start, see Reid (1778/1969), Chisholm (1966), Clarke (1993; 2003), O'Connor (2000; 2005; 2009), and Pereboom (2001; 2014), especially these last three authors.

negative claim: it claims that substances or agents cause events, but in a way that *can't* be reduced to or grounded in events causing events. But if one isn't clear on what substance- or agent-causation is to begin with, this negative characterization doesn't seem to help.

This worry for substance- and agent-causation has been around for some time, and many authors have sought to address it.[7] Unfortunately, evaluating this worry is far beyond the scope of this chapter. For our purposes, it is sufficient to note that the jury is still out on the intelligibility of substance- and agent-causation. For the libertarian, it would be preferable to have a solution to the problem of enhanced control that didn't hinge on such controversial metaphysics.

There is another worry for the agent-causal solution apart from its intelligibility. Recall claim (ii): that agent-causation requires the falsity of determinism. This claim is dubious. Determinism would seem to be compatible with the existence of substances and agents. Why, then, would determinism rule out substance- or agent-causation? Several authors who accept the general notion of substance- or agent-causation even explicitly state that it is *compatible* with determinism.[8] Absent some compelling argument, it is unclear why we should think otherwise.[9]

[7] I find O'Connor (2000), Clarke (2003), and Pereboom (2004) among the most illuminating.

[8] Markosian (1999; 2012), Clarke (2003), and Nelkin (2011) give agent-causal accounts that are compatible with determinism. See Franklin (2016) as well, although Franklin does not accept an agent-causal account nor compatibilism.

[9] The only author I know of who gives an explicit argument for (ii) is Tim O'Connor (2003; 2009). Suffice it to say that not many have been convinced.

Allow me to clarify the dialectic here. I am *not* claiming that the notion of substance- or agent-causation is unintelligible, nor am I claiming that substance- or agent-causation is compatible with determinism. Rather, I am simply pointing out that substance- or agent-causation isn't *obviously* an intelligible notion, and that it isn't *obviously* incompatible with determinism. This means that, for the libertarian, the promise of the agent-causal solution to the problem of enhanced control comes with potential costs: it comes with the burden of defending a mysterious notion as well as showing that this mysterious notion is incompatible with determinism. All else being equal, an answer that had the promise of the agent-causal solution, but didn't take on these burdens, would be preferable. Shortly, I'll argue that my proposed solution to the problem of enhanced control has exactly these features.

Franklin's Solution

Christopher Franklin (2011; 2018) has given a very different solution to the problem of enhanced control. He claims that indeterminism (in the right places) enhances our control by giving us *opportunities* to exercise our abilities. For instance, suppose a pianist is stranded in the middle of the desert for an afternoon with no piano around for hundreds of miles. While the pianist retains the *ability* to play the piano, her environment does not afford her the *opportunity* to exercise that ability. And because of that, she doesn't have as much control over how she spends her afternoon as she otherwise would. Franklin says something similar holds of the thief. In the deterministic version, he retains the *ability* to refrain from stealing, but determinism denies him the *opportunity* to do so; in the indeterministic version, he not only has the ability to refrain from stealing but also the opportunity, thereby giving him more control over how he behaves in the church.

As with the agent-causal solution, I think there is something to be said in favor of Franklin's solution. First, Franklin's solution doesn't seem to require any mysterious notion, certainly not substance- or agent-causation, although it is consistent with such notions. Second, Franklin explicitly gives an appealing argument for the claim that opportunities are incompatible with determinism. He starts with the following:

> *S* has an opportunity to φ only if there is no decisive obstacle to his φ-ing—only if nothing prevents him from φ-ing. An agent's failure to φ does not itself constitute a decisive obstacle to his φ-ing… But it would seem that if his performing the action required any more difference than this, he would lack the opportunity. If, in addition to his φ-ing, something in his environment must be different—a difference that would not be identical to or dependent on his φ-ing—then it seems that this required difference is a decisive obstacle: it prevents him from doing otherwise. It prevents him from doing otherwise since the only worlds in which he φs differ from the actual

> world, and these differences are neither an exercise of, nor dependent on an exercise of, his agency. (2018, pp. 71-72)

According to Franklin, *S* has the opportunity to φ only if there is no decisive obstacle to *S*'s φ-ing, and there is no decisive obstacle only if *S*'s φ-ing does not require a difference in *S*'s environment that is independent of *S*'s φ-ing. But if determinism is true and *S* doesn't φ, *S*'s φ-ing requires that either the past or the laws be different. And, as Franklin later suggests, since the past and the laws are independent of *S*'s φ-ing, it follows that determinism presents a decisive obstacle to *S*'s φ-ing, which means determinism robs *S* of the opportunity to φ. More generally, if determinism is true, then no one ever has the opportunity to do otherwise.

If this argument sounds somewhat familiar, it should, as I gave a very similar argument in the last chapter. The argument went like this (roughly): *S* is free to perform action *X* at time *t* only if it is possible for *S* to perform *X* at *t* holding fixed all of those facts that are distinct from and explanatorily independent of *S*'s behavior at *t*. But the laws and the past are distinct from and explanatorily independent of *S*'s behavior at *t*. So, if determinism is true and *S* doesn't perform *X* at *t*, then *S* isn't free to perform *X* at *t*. Franklin's argument is focused on opportunities rather than freedom itself, and it's not obvious he is utilizing *explanatory* independence, but the two arguments are siblings at least, if not twins. Hence, by my lights, Franklin has an appealing argument for the claim that the opportunity to do otherwise is incompatible with determinism.

Despite these advantages, though, Franklin's solution still leaves something to be desired. The central issue is that it seems to simply push the problem of enhanced control back a step. When we first formulated the problem of enhanced control, we put it like this: "According to libertarianism, indeterminism allows for the freedom to do otherwise, which means that indeterminism enhances one's control. But how does that work?" Under Franklin's proposal, it would seem we could rerun the problem like this: "According to (Franklin's) libertarianism, indeterminism allows for opportunities to do otherwise, which means indeterminism enhances one's control. But how does that work?" This question, although different in some ways, still deserves an answer. While Franklin's proposal may *focus* the problem of enhanced control, it's not so clear that it *solves* it.

Here's a way to see this point. As noted, the original formulation of the problem of enhanced control can grant *that* the freedom to do otherwise is incompatible with determinism. But merely showing *that* determinism precludes freedom doesn't show *how* indeterminism secures this freedom, and that's what generates the problem. Now with Franklin, we shift from the *freedom* to do otherwise to the *opportunity* to do otherwise, but it's not clear that this shift helps answer the problem in a fully satisfactory way. It would seem that the problem of enhanced control can grant *that* the opportunity to do otherwise is incompatible with determinism. But merely showing *that* determinism precludes opportunities doesn't show *how* indeterminism secures these opportunities, and so a problem still remains. Just think of the thief's case again: even if determinism *takes away* the thief's opportunity to refrain from stealing, it isn't immediately clear how adding indeterminism – a little bit of randomness – somehow *gives* the thief this opportunity. For all that Franklin says, it could turn out that indeterminism also takes away the thief's opportunity.

There is a more general takeaway here: in order to give a fully satisfactory answer to the problem of enhanced control, we must invoke some notion that is not only *precluded* by determinism, but clearly *granted* by indeterminism. Franklin has helped us see that the freedom to do otherwise requires both the *ability* and the *opportunity* to do otherwise, and that determinism rules out the latter. That is genuine progress. But it isn't immediately obvious how indeterminism *grants us* the opportunity to do otherwise, which is what allows the problem of enhanced control to be rerun against Franklin's solution. Shortly, I'll argue that my proposed solution avoids this trouble because the notion in question is not only ruled out by determinism, but clearly granted by indeterminism.

Summing Up

The primary purpose of this section has been to argue that the extant solutions to the problem of enhanced control are unsatisfactory. It's not, necessarily, that the proposed solutions are incorrect or mistaken; it's that they suffer from some dialectical problems. Thinking through these shortcomings suggests three desiderata for any solution: (1) the solution avoids controversial notions; (2) the solution utilizes a notion that is plausibly precluded by determinism; and (3) the solution invokes a notion that is clearly granted by indeterminism, so as to make sure the problem of enhanced control cannot simply be rerun. I now move to my proposed solution which I will argue meets all of these desiderata.

Eliminable and Ineliminable Explanations

Just as the agent-causal solution claims that indeterminism allows agents to play a certain kind of *causal* role, so my solution claims that indeterminism allows agents to play a certain kind of *explanatory* role, what I will call an *ineliminable* part of an explanation. To get there, we'll first start with the idea of an *eliminable* part of an explanation and a simple example.

Suppose there are 1,000 dominos configured in such a way that, for all natural numbers, *n*, if the nth domino falls over, then it will knock over the nth+1 (assuming there is an nth+1 domino), and suppose the first domino falls over for some inscrutable reason, thereby knocking the second domino over, which knocks over the third, and so on. Now ask this question: why did the 1,000th domino fall over? The best available explanation, it seems, is this:

> **Explanation 1**: (i) The dominos are configured in such a way that if the nth domino falls over, then the nth+1 domino will fall over, and (ii) the first domino fell over.

This explanation has all that you could ask for. It is simple and straightforward, makes the 1,000th domino's falling over as likely as possible, and explains just about everything there is to be explained in this context. Now compare this explanation to the following:

> **Explanation 2**: (i) The dominos are configured in such a way that if the n^{th} domino falls over, then the $n^{th}+1$ domino will fall over, and (iii) the 501st domino fell over.

This isn't a bad explanation by any means. It's not even a rival to **Explanation 1** – it is in some sense a *part* of **Explanation 1**. Moreover, in some cases we might prefer **Explanation 2** to **Explanation 1** for pragmatic reasons. For instance, maybe someone is unaware of dominos 1 through 500 and you don't want to have to inform them about the extra dominos. Or maybe you are talking to a young child who has just learned about quantities in the hundreds but hasn't dreamt of quantities in the thousands. In such cases, **Explanation 1** might be less helpful than **Explanation 2**. Nevertheless, I hope you will agree that **Explanation 1** is, in an important sense, a better explanation than **Explanation 2**. After all, **Explanation 1** not only explains everything that **Explanation 2** does, but it explains lots of other things as well, such as why the 501st domino and all the ones before it (other than the first) fell over.

It will be helpful to introduce some terminology. Let's say that whether one explanation is better than another is determined by how *well* it explains the phenomena in question (its "predictive power") as well as how *many* related phenomena it can explain (its "scope").[10] We can say that **Explanation 1** is better than **Explanation 2** because it not only explains everything just as well as **Explanation 2** does, but it also explains more than **Explanation 2** does – that is, **Explanation 1** has at least as much *predictive power*, and certainly a wider *scope*, than **Explanation 2**.

Finally, consider one last explanation we could give:

> **Explanation 3**: (i) The dominos are configured in such a way that if the n^{th} domino falls over, then the $n^{th}+1$ domino will fall over, (ii) the first domino fell over, and (iii) the 501st domino fell over.

This is a funny explanation. Fact (iii) seems wholly redundant. We can "derive" fact (iii) from (i) and (ii) (and the fact that there are at least 501 dominos). And not just that, but we can also fully *explain* fact (iii) using facts (i) and (ii). Intuitively, we can just "delete" fact (iii) from the explanation without any loss in either predictive power or scope: an explanation for why the 1,000th domino fell over isn't necessarily any worse off for not including the fact that the 501st domino fell over.

To capture this, let's say that fact (iii) is an *eliminable* part of the explanation for why the 1000th domino fell over. More precisely, let's say that a fact

[10] A natural way to understand how *well* an explanation explains the things it does is in terms of probability: an explanation explains some phenomenon perfectly well only if the probability of the phenomenon occurring given the explanation is 1; any explanation on which the probability of the phenomenon occurring is less than 1 explains the phenomenon less well.

(or set of facts), *P*, is an eliminable part of the explanation of some other fact (or set of facts), *Q*, if the following conditions hold:

1. *P* is at least part of the explanation of *Q*.
2. There is some distinct, non-overlapping fact (or set of facts), *R*, such that *R* is at least part of the explanation of *Q*.
3. *P&R* is no better an explanation of *Q* than just *R*.[11]

Intuitively, the definition is claiming that, although *P* is a part of the explanation of *Q*, adding it to another set of facts, *R*, doesn't add anything with regards to either predictive power or scope. In the case at hand, the fact that the 501st domino fell over meets conditions (1), (2), and (3): the fact that the 501st domino fell over partly explains why the 1,000th domino fell over, thereby satisfying condition (1); the facts that the 1st domino fell over and that the dominos are arranged in the relevant way also explain why the 1,000th domino fell over (and don't overlap with the fact that the 501st domino fell over), thereby satisfying condition (2); and the conjunction of these three facts is no better an explanation than just the conjunction of these last two facts, thereby satisfying condition (3). So, the fact that the 501st domino fell over is an eliminable part of the explanation for why the 1,000th domino fell over.

[11] This definition might have some counterintuitive implications for cases of overdetermination. If *P* fully explains *Q*, and some wholly distinct set of facts, *R*, also fully explains *Q*, then this definition implies that both *P* and *R* are eliminable parts of the explanation for *Q*. I'm not sure if that's a "bug" or a "feature" of the definition. Perhaps the thing to say is that, while both *P* and *R* are individually eliminable, the disjunction of *P or R* is not. Or perhaps we should slightly amend (3) so that *R* also entails *P*. Either way, I'll put aside the issue of overdetermination for now. Thanks to Taylor Cyr for help here.

Now that we have a grip on an *eliminable* part of an explanation, we can state what constitutes an *ineliminable* part of an explanation. Most simply, an ineliminable part of the explanation of *Q* just is a part of the explanation of *Q* that isn't eliminable. To put it more formally, *P* is an ineliminable part of the explanation of *Q* if:

4. *P* is at least part of the explanation of *Q*.
5. There is no distinct, non-overlapping fact (or set of facts), *R*, such that *P&R* is no better an explanation of *Q* than just *R*.

Intuitively, an ineliminable part of an explanation is some fact that *must* be cited if the explanation is to have a maximal amount of predictive power and scope. For instance, in our domino case, the fact that the first domino fell over is an ineliminable part of the explanation for why the 1,000th domino fell over. If an explanation doesn't cite that fact, then the explanation is necessarily worse off for it.

Is there a general rule for when a fact will qualify as an ineliminable part of an explanation? The beginnings of causal chains will usually (if not always) count as ineliminable parts of explanations, as will the fundamental laws governing such chains. It is even tempting to think that *only* these kinds of facts will qualify as ineliminable parts of explanations. But that's too quick. In particular, and most relevant for our purposes, *indeterminism* in the right places can allow for facts further down the causal chain to count as ineliminable parts of an explanation.

Consider a variant on our simple domino case again. Everything is exactly the same except for one miniscule difference: if the nth domino falls over, then the nth+1 domino will fall over, *the exception being* that if the 500th domino falls over, then it is only *more likely than not* that the 501st domino will fall over. (Maybe the dominos have been setup such that the 500th domino has to fall off of a table and bounce a certain way in order to knock over the 501st domino.) To give it a number, let's say that if the 500th domino falls over, there is an objective probability of 0.7 that the 501st domino will fall over. Otherwise, everything is the same.

Just as before, the first domino falls over for some inscrutable reason, knocking the second, which knocks the third, and so on until the 1,000th domino is knocked over. Now ask the same question: why did the 1,000th domino fall over? Compare variations on our explanations again:

> **Explanation 1***: (i) The dominos are configured in such a way that if the nth domino falls over, then the nth+1 domino will fall over, with the exception of the 500th and 501st dominos, and (ii) the first domino fell over.

> **Explanation 2***: (i) The dominos are configured in such a way that if the nth domino falls over, then the nth+1 domino will fall over, with the exception of the 500th and 501st dominos, and (iii) the 501st domino fell over.

> **Explanation 3***: (i) The dominos are configured in such a way that if the nth domino falls over, then the nth+1 domino will fall over, with the exception of the 500th and 501st dominos, (ii) the first domino fell over, and (iii) the 501st domino fell over.

By adding indeterminism in the right place, fact (iii) now qualifies as an ineliminable part of the explanation for the 1,000th domino's falling over. To see this, notice that **Explanation 3*** is a better explanation than both **Explanation 1*** and **Explanation 2***. Why? Start with **Explanation 1***. Clearly, **Explanation 1*** and **Explanation 3*** have the same scope – there isn't any fact that one explanation can explain but the other can't. But **Explanation 3*** has the advantage with regards to predictive power. In particular, the likelihood of the 1,000th domino falling over is only 0.7 under **Explanation 1*** whereas it is 1 under **Explanation 3***. Hence, **Explanation 3*** is superior to **Explanation 1*** because it has just as much scope but more predictive power.

Now consider **Explanation 2***. **Explanation 2*** and **Explanation 3*** seem to make the 1,000th domino's falling over equally likely, but **Explanation 3*** has the resources to explain something that **Explanation 2*** doesn't, namely, why the 501st domino fell over. In **Explanation 2***, the 501st domino's falling over is in no way explained – **Explanation 2*** doesn't even *attempt* to explain that. But **Explanation 3*** can offer some explanation for why the 501st domino fell over: the first domino fell over, which caused the second, which caused the third, etc., until the 500th fell over, which *likely* caused the 501st to fall over. It's not a *full* explanation perhaps, but it's better than nothing.[12] Hence, **Explanation 3*** is superior to **Explanation 2*** not because it has more predictive power with regard to the 1,000th domino falling over, but because it has a wider scope.

More generally, there seems to be no fact or set of facts, *R*, such that (iii) in conjunction with *R* is no better an explanation than just *R*: if an explanation does not include the fact that the 501st domino fell over, it will necessarily lose out on predictive power or scope. So, it is not just beginnings of causal chains that qualify as ineliminable parts of an explanation. By adding indeterminism in the right places, facts much further down the causal chain can also qualify.

With a grip on the notion of an ineliminable part of an explanation and its relation to indeterminism, we can return to the problem of enhanced control.

The Explanatory Solution and the Three Desiderata

I'm sure the reader can anticipate where we are headed. As we've seen, the problem of enhanced control centers around this question: "Even if determinism rules out a certain kind or degree of control, how is indeterminism supposed to help?" The agent-causal solution answers: "Indeterminism (in the right places) allows for us to play a certain kind of causal role – it allows us to be agent-causes of our

[12] Carolina Sartorio has pointed out to me that, if we assume any given domino falls over only if knocked by the previous one (assuming there is a previous one), perhaps **Explanation 3*** (or something quite similar) can give a full explanation. If so, then **Explanation 3*** is far superior to **Explanation 2*** with regard to scope.

decisions, deliberations, etc.” Franklin answers: “Indeterminism (in the right places) gives us the opportunity to exercise certain abilities – it gives us the opportunity to do otherwise.” I instead answer: “Indeterminism (in the right places) allows us to play a certain explanatory role – it allows our decisions, deliberations, etc. to be *ineliminable* parts of the explanation for how the world unfolds, thereby increasing the significance of our agency.” I’ll call this the “explanatory solution.”

Let’s return to the thief’s case and add a bit to the story. Suppose that the church funds a local shelter, but money is becoming quite scarce. Without a big influx in giving, the church will have to stop supporting the shelter, which means the shelter will close down. And sure enough, the chunk of change the thief steals makes the difference: the church just barely misses the requisite funds, and so the shelter closes its doors. For simplicity, suppose the connection between the thief’s decision and the shelter’s closing is deterministic, regardless of whether the thief’s decision is brought about in a deterministic way or not.

What explains why the shelter shut down? In the fully deterministic version of the story, the thief’s decision to steal the money is plainly part of the explanation, and an important part in some sense. But the thief’s decision is an *eliminable* part of the explanation: in principle, one need not cite the thief’s decision at all, but can merely cite the initial conditions of the universe and the laws. If determinism is true, then citing just the initial conditions of the universe and the laws, but not the thief’s decision, will have at least as much predictive power and scope as any explanation which also cites the thief's decision. Of course, explanations which cite the thief's decision will often be preferable for pragmatic reasons. If someone were to ask why the shelter shut down, it would be borderline sadistic to start with "Well, everything started with the Big Bang and a cosmic soup of quarks, electrons, bosons, and…" But this is irrelevant to whether the thief's decision is ineliminable or not. Recall our first domino case: there may very well be some contexts where explanations which cite the 501st domino's falling over are preferable for pragmatic reasons. Still, so long as the 501st domino's falling over doesn't necessarily add to an explanation's predictive power or scope, it remains eliminable.

In the indeterministic version, where there was only an objective probability of 0.7 that the thief would decide to steal, matters are much different. It would seem that, by introducing a little indeterminism, the thief's decision now becomes an *ineliminable* part of the explanation for why the shelter closed its doors: any explanation which didn't cite the thief's decision would necessarily lose out on predictive power or scope. For instance, an explanation which only cited the initial conditions and the laws would have plenty of scope, but it would miss out on some predictive power – it would only make the shelter's closing objectively probable to a degree of 0.7. In order for an explanation to have a maximal amount of predictive power and scope, it needs to cite the thief's decision. That implies that the little bit of indeterminism allows for the thief's decision to be an ineliminable part of the explanation for the shelter's closing.

Granted, the thief's case is greatly simplified. But the general point remains: indeterminism (in the right places) allows us to play a certain explanatory role – it allows our values, desires, decisions, etc. to be *ineliminable* parts of the explanation for certain events in the world. If our universe is deterministic, then everything we ever do only plays a redundant or eliminable explanatory role: our values, desires, decisions, etc. add nothing whatsoever to the best explanation for any given event. But if indeterminism holds (in the right places), then the best explanation for certain events necessarily cites some facts about us, facts about our values, desires, decisions, etc. To put it in my grandiose way, if indeterminism holds (in the right places), then one cannot tell, or at least explain, the history of the world without citing at least some of the facts of our agency.

So that is the explanatory solution. What should we make of it? Let's start by considering those three desiderata proposed earlier: (1) the answer avoids controversial notions; (2) the answer uses some notion that is plausibly precluded by determinism; and (3) the answer uses a notion that is clearly granted by indeterminism, so as to ensure that the problem of enhanced control cannot simply be rerun. I'll take each in turn, arguing that the explanatory solution meets them all quite well. I'll then close by considering an objection unique to the explanatory solution.

The First Desideratum: Avoiding Controversial Posits

In contrast to agent-causal solutions, the explanatory solution is consistent with, but does not require, any terribly controversial posits. The explanatory solution only requires that *some* relevant feature of agency play an ineliminable explanatory role, whether that be certain agential events – decisions, tryings, deliberations, etc. – or the agent herself (full-stop). To return to our thief case, the explanatory solution is consistent with the thief agent-causing his decision, as that may very well imply that the thief plays an ineliminable explanatory role in the shelter's closing down. But the explanatory solution does not require that the thief agent-cause his decision – there are many other ways the thief's agency could play an ineliminable explanatory role. So long as the thief's agency plays such a role, the explanatory solution is satisfied.

The explanatory solution does require thinking of explanation in a certain way, namely, that some explanations are objectively better than others in virtue of having more predictive power or scope. Suspicions may arise here from those with strong pragmatist leanings, but this issue is a bit beyond the scope of this chapter. Suffice it to say that the nature of explanation I have presupposed here is far less controversial than anything like the notion of substance- or agent-causation. So, at the very least, the explanatory solution seems to require far less controversial posits or assumptions.

The Second Desideratum: Being Precluded by Determinism

The relationship between the explanatory solution and determinism is more complicated. Determinism doesn't necessarily rule out an agent playing an ineliminable explanatory role; it only rules out *agents sufficiently similar to us* playing an ineliminable explanatory role. But far from being a problem, this turns out to be exactly the right result, or so I will argue.

Initially it might look as if, given the truth of determinism, only the initial conditions and the laws governing those conditions will qualify as ineliminable parts of an explanation for any given event. So, it would seem to follow that determinism rules out any agent playing an ineliminable explanatory role with regard to any given event. But this is mistaken. It's not too difficult to imagine cases where an agent plays an ineliminable explanatory role despite being in a deterministic world. Here are three.

First, consider a case given by Joseph Campbell:

> Suppose that [world] *W* is a determined world such that some adult person exists at every instant… At its first moment of existence lived Adam, an adult person with all the knowledge, powers, and abilities necessary for moral responsibility. (2007, p. 109)

Now Campbell isn't concerned with the problem of enhanced control or ineliminable explanations. Instead, Campbell takes this case to show that it is only contingently true that no one has ever had a choice about the remote past, like the initial conditions. But we can use the case for our own purposes. Since Adam is part of the initial conditions of *W*, Adam can still play an ineliminable explanatory role despite the fact that *W* is deterministic. So if this is a genuinely possible case, determinism doesn't necessarily rule out an agent's playing an ineliminable explanatory role.

We can also imagine cases where an agent is not part of the initial conditions but still might play an ineliminable explanatory role with respect to those conditions. Consider a case from Carolina Sartorio:

> Time-Traveler Adam lives in a world where time-travel is physically possible and he is in fact in possession of a working time machine that could take him to any time, including any past time… Although he could have traveled to those past times, he didn't in fact travel to those times. (2015, p. 260)

Sartorio, like Campbell, uses this case to show that it is possible for an agent to have a choice about the remote past. But again, it also has implications for the explanatory solution. Importantly, Time-Traveler Adam does *not* exist at the first moment of this deterministic world, although he *could have* if he had decided to use his time machine. And this suggests that, although he is not a part of the initial conditions, facts about his agency partly explain the obtaining of those initial conditions – after all, if he had used the time machine to travel to the first moment, the initial conditions might have been different. This suggests that an explanation of the world which only cited the initial conditions might not have a maximal amount of scope which, in turn, may allow for Time-Traveler Adam to play an ineliminable explanatory role, even though his world is deterministic. More generally, if facts about an agent somehow explain the initial conditions, it's not obvious that determinism disqualifies the agent from playing an ineliminable explanatory role.[13]

Finally, consider an agent that has the power to break the (actual) laws of physics, but never exercises that power. Here's a particular example:

> Miracle-Worker Adam is a young agent living in an old and deterministic world, but has been given a great gift: should he decide, he could suspend the law of gravity and have objects float away from the surface of the Earth. He likes things the way they are though, and never uses this power.[14]

Unlike Adam and Time-Traveler Adam, Miracle-Worker Adam never has access to the initial conditions, meaning he does not explain the obtaining of those

[13] The case is quite tricky, though, because it seems to involve an explanatory loop: although Adam's decision to not use the time machine explains the initial conditions, those very same initial conditions seem to explain (at least ancestrally) his decision, given the truth of determinism. It might seem, then, that an explanation which cited just the initial conditions would have just as much predictive power and scope as one which instead cited the state of the world at the time of Adam's decision, which would imply that both the initial conditions and Adam's decision are *eliminable*! But at the very least, there wouldn't be a *better* explanation of the world which didn't invoke Adam's decision. That seems like an important difference between Adam's decision and our decisions (if determinism holds), one we could capture by slightly amending the definition of an "eliminable" explanation.

[14] Brian Cutter (2017) also considers such examples, although they are a bit weaker. (His "coy miracle-workers" are able to perform some action such that, if they were to perform it, an actual law would have been violated, although not necessarily by the agent's action.) But Cutter is focused again on a different issue, namely, the claim that, necessarily, no one has ever had a choice about the laws. Thanks to Taylor Cyr for pointing this out.

conditions. Nonetheless, he does seem to have some explanatory relation to the laws governing those conditions: it seems like his choices at least partly explain why the law of gravity holds. For parallel reasons, this suggests that Miracle-Worker Adam may play an ineliminable explanatory role with respect to certain events in his world, even though his world is deterministic.

All three of these cases are, of course, quite controversial. But even if they are impossible, they raise an important point: it's not obvious that determinism *itself* precludes any agent from playing an ineliminable explanatory role. It is only determinism *for agents like us* – agents that have never had access to the initial conditions nor power over the laws – that determinism poses a threat for. And this lines up quite nicely with how others have suggested we think about the problem of freedom and determinism. I can do no better than to quote Sartorio again:

> The problem [of freedom and determinism] doesn't arise, as we had thought, only because determinism allegedly *results in* the absence of an important form of control. Instead, it arises insofar as determinism is *conjoined with* the absence of control (as it happens in our case). In other words, the lack of control is not only an alleged implication of determinism, but it is also part of what gives rise to the problem. It is an important *source* of the problem. (2015, p. 261; emphasis in original.)

According to Sartorio, the mere fact that our behavior is determined by the past and the laws doesn't pose a problem for freedom. Rather, it is our behavior being entailed by the past and the laws, *neither of which we have control over,* that (allegedly) poses a problem for freedom.[15] If we had control over the past or the laws – if we had a time machine or were miracle-workers – then determinism wouldn't obviously threaten our freedom.

This fits just about perfectly with the explanatory solution. It is not determinism *itself* that precludes an agent from playing an ineliminable explanatory role. Rather, determinism precludes *agents like us* from playing an ineliminable explanatory role. So it looks like the explanatory solution meets the second desideratum in a deeply satisfactory way.

The Third Desideratum: Not Relocating the Problem

In contrast to Franklin's solution, the explanatory solution does not seem to simply relocate the problem of enhanced control. Recall that according to Franklin's solution, indeterminism enhances an agent's control by giving the agent more *opportunities* to exercise her abilities – only under indeterminism does the agent ever have the opportunity to do otherwise. The worry is that the problem of enhanced control can be rerun. Just as it is unclear how indeterminism enhances one's control, so it is unclear how it enhances one's opportunities. As it was put earlier, even if determinism rules out certain opportunities, how is it that

[15] To be clear, Sartorio holds that the freedom required for moral responsibility is compatible with determinism, even if the freedom to do otherwise is incompatible with determinism. I wish to remain neutral on the relationship between the freedom to do otherwise and moral responsibility.

introducing a certain kind of *randomness* gives the agent those opportunities back?

The explanatory solution avoids this issue. If we were to rerun the problem of enhanced control against the explanatory solution, it would go something like this: how is it that indeterminism changes an agent's explanatory role? How is it that introducing a certain kind of *randomness* allows the agent to play an ineliminable explanatory role? These questions have straightforward answers. Think about our domino cases again. Not only does determinism preclude the 501st domino from playing an ineliminable explanatory role, but indeterminism – a certain kind of randomness – straightforwardly allows it to play such a role. That's because randomness allows for facts further down the causal chain to contribute to an explanation's predictive power. Therefore, it is not only plausible *that* determinism would rule out our playing an ineliminable explanatory role, but it is also clear *how* indeterminism would let us play such a role. If so, then the problem of enhanced control cannot be rerun.

This is no coincidence. Franklin's solution relies on a notion that is extremely close to the notion of freedom – it would seem that any initial *analysis* of freedom, libertarian or otherwise, should invoke the notion of an opportunity. Given this proximity, it's not terribly surprising that the problem of enhanced control can simply be rerun. By comparison, the notion of an ineliminable explanatory role is quite distant from freedom. For instance, whereas it is unclear whether non-agents can *literally* have opportunities, they can certainly play ineliminable explanatory roles. At the very least, the notion of an ineliminable explanation is not in the initial analysis of freedom, whereas the notion of an opportunity arguably is. It is for this reason, I suspect, that the problem of enhanced control cannot simply be rerun against the explanatory solution.

But the distance between the notions of freedom and ineliminable explanations is a bit of a double-edged sword, for it also gives rise to a problem unique to the explanatory solution. Some might argue that the distance is simply *too* great: it's not clear, some might suggest, how the notion of an ineliminable explanatory role is *related to freedom at all*. Imagine someone saying the following:

> Fair enough, indeterminism (in the right places) allows us to play an ineliminable explanatory role. But *who cares*? What does playing such a role have to do with *freedom* or *control*? I can see why we would want to play *some* explanatory role, or even an *important* explanatory role with respect to various events in the world. But you've admitted that the truth or falsity of determinism isn't relevant here. What appeal does playing an *ineliminable* explanatory role have over and above these other explanatory roles?

This is indeed a significant challenge, one that arguably any solution which meets the third desiderata will have to face. I'll spend the entirety of the next section addressing it.

Ineliminable Explanation and Authorship

Most basically, I think we should care about playing an ineliminable explanatory role because it allows for us to be the *author* of certain events in a deep and important way. Or at least the libertarian should say as much. To see the connection between ineliminable explanations and authorship, let's consider two cases. We'll start by considering a case where the agent does *not* play an ineliminable explanatory role; then we'll consider a case where the agent *does* play an ineliminable explanatory role. The difference between these cases lines up with a difference in *authorship* as well which, I'll argue, gives us good reason to think the two notions are intertwined.

The first case is taken from the film *Stranger than Fiction*. Will Ferrell plays an IRS agent, Harold, who leads a largely unremarkable life (if one can imagine that). But one feature of Harold's life is extremely peculiar: it is literally *being written* by someone else. In some instances, he can even hear the author, Karen (played by Emma Thompson), *narrating* his life, making infallible predictions as to what is about to happen to him. Eventually, he discovers that the author is planning to kill him off in the near future, which causes him to go looking for this mysterious narrator with the goal of convincing her to spare him.[16]

[16] This case bears important similarities to cases often invoked in so-called "manipulation" arguments, although such arguments focus on the notion of moral responsibility rather than authorship. See Mele (2006), Fischer (2011), Kearns (2012), and Todd (2013), for a start.

For simplicity, let's modify the case just a bit. Suppose instead that Harold never finds out about Karen nor about her plans to kill him off. As far as he is concerned, he is the sole author of his life. He is mistaken, of course, but never learns this. Further, suppose that Karen is both an incredibly meticulous writer, having authored nearly *every event* of his life, and also a very good writer, making sure Harold and his life exhibit standard human qualities.

Arguably, Harold isn't the *author* of his actions or of his life in general – he is *literally* just living out a story written by someone else! But it's not immediately clear *why* he isn't the author. What more could he possibly need? After all, he seems to fulfill all of the standard requirements for a robust sense of agency: he is responsive to reasons, takes ownership of his actions, has deeply held values, lacks significant conflict between his lower-order and higher-order judgments, etc.[17] (Or at the very least we can imagine a case like this.) The only difference between us and Harold seems to be that *someone else* is writing his story, whereas we typically think of *ourselves* as "writing the story of our life." But how, exactly, does this make a difference? Can we give a more general account?

The notion of an ineliminable explanation can provide an answer. Given the setup of the case, it appears as if Harold's decisions, values, deliberations, etc. only play an *eliminable* explanatory role with regard to the overall shape of his life. For

[17] These conditions are sometimes called standard "compatibilist" conditions, although they are typically given with respect to moral responsibility, not authorship. See Frankfurt (1971), Fischer & Ravizza (1998), Watson (2004), McKenna (2012), and Sartorio (2016) for just the tip of the iceberg.

instance, if we want to know why Harold is an IRS agent, we of course *could* cite certain of his decisions, values, and deliberations. But we don't *need* to. We could instead merely cite Karen's vision for his life and her decision to make him an IRS agent. With the connection between Karen's decisions and Harold's life being as strong as it is, an explanation that only cited facts about Karen's agency would be just as good of an explanation, if not better, than one which cited facts about Harold's agency. So, if being the author of one's life requires facts about one's agency to play an ineliminable explanatory role with respect to one's life, we have a deep explanation for why Harold isn't the author of his life.

So Harold's case seems to be one where the agent is neither the author of, nor plays an ineliminable explanatory role with respect to, the overall shape of his life. But let's now focus on a case where the agent is both the author and plays an ineliminable explanatory role, one that is particularly puzzling for the libertarian.

Consider a traditional western conception of God according to which God is *essentially* all-powerful, all-knowing, and morally perfect: that, necessarily, God has the power, knowledge, and will to perform the morally best action in any given circumstance (supposing there is a best action). This conception of God has gotten libertarians into knots for a long time. On the one hand, such a being seems to be the *paradigm* of authorship. Having no external limitations, as well as perfectly formed desires and judgments, God would seem to be the author not just of her own life, but possibly *everyone's* life. But on the other hand, such a being does not seem to be *free*, at least not by the libertarian's lights. Since God is *essentially* all-powerful, all-knowing, and morally perfect, God *cannot* do anything but the best action, meaning God is never free to do otherwise. While God may not be determined by the laws of physics and initial conditions, she is nonetheless determined by her character, one which she has of necessity.[18] How, then, can God be seen as the author of any event, especially by the libertarian?

Tying authorship with ineliminable explanation can do the trick. Suppose God brings about some event, like the creation of the world. Sure enough, God may not have been able to do otherwise. But focus instead on a different question: why is it that the world was created? What explains this event? The best explanation would seem to cite various features of the world, like it's being the best possible world perhaps, as well as certain facts about God, like the fact that God has the desire and power to bring the best about. Any explanation which didn't cite such facts would necessarily lose out on either predictive power or scope – the best explanation *necessarily* cites facts about God's agency. Hence, it seems as if facts about God's agency are *ineliminable* parts of the explanation for why the world was created. If authorship requires, first and foremost, playing an ineliminable explanatory role, the libertarian has no reason to think that God isn't the author of the world.

When taken together, these two cases give us strong (albeit defeasible) reason for thinking that authorship of an event requires facts of one's agency to

[18] Howard-Snyder (2009) pushes this type of worry, concluding that God so understood is not praiseworthy for any action under standard libertarian (or incompatibilist) assumptions.

play an ineliminable explanatory role with respect to that event. Or at the very least, the libertarian should take this connection seriously. If so, then the libertarian ought to regard the explanatory solution as the most plausible response to the problem of enhanced control.

Concluding Remarks: Toward a Unified Incompatibilism

In short, the explanatory solution claims that indeterminism enhances our agency because it allows us to play an ineliminable explanatory role in the unfolding of the world – that the story of the world cannot be explained without citing various features of our agency. The explanatory solution has a number of appealing features: it invokes relatively uncontroversial notions, ones that are incompatible with determinism (at least for agents like us), but are granted under indeterminism, thereby making sure the problem of enhanced control cannot simply be rerun. So long as the notion of an ineliminable explanatory role is *somehow* connected to an agential notion – such as the notion of *authorship*, as suggested above – the explanatory solution would appear to be the best answer to the most pressing challenge for libertarianism.

But before I let you go, there is one more feature of the explanatory solution worth noting. Those who accept that determinism is incompatible with freedom come in two varieties. According to so-called "leeway" incompatibilists, determinism would rule out our free agency because it would rule out our being free to do otherwise. The libertarianism discussed in this chapter is most naturally interpreted as a kind of leeway incompatibilism. According to so-called "source" incompatibilists though, determinism would rule out our free agency because it would rule out our being the "source" of our actions or various events in the world.[19] While these two types of views are often seen as quite distinct, the explanatory solution, in conjunction with the comments above about authorship, promises to unify them in a deep way.

Most basically, the suggestion is this: we care most centrally about playing an ineliminable explanatory role with respect to the happenings in the world, which is right in line with source incompatibilism. But it turns out that, in order for *agents like us* to play an ineliminable explanatory role, we must have the freedom to do otherwise. Perhaps there are some possible agents, such as Adam or God, that can play ineliminable explanatory roles without the freedom to do otherwise. But for agents that have external constraints and no control over the past or the laws, the freedom to do otherwise – and hence indeterminism – is needed in order to play such a role. If so, then we can truthfully say *both* that determinism would rule out our free agency because it would rule out our being free to do otherwise *and* because it would rule out our being the "source" of our actions. By using the explanatory solution, we find a version of incompatibilism

[19] This distinction is originally from Pereboom (2001; 2014), although many others have since taken it up.

that seems to respect both of the intuitions driving leeway and source incompatibilism.[20]

Of course, this unification deserves further development. For instance, leeway incompatibilists don't care about any old alternative course of action, but rather "robust" alternatives.[21] It's far from obvious how to connect the notion of an ineliminable explanation with a "robust" alternative course of action. On the other side, it's unclear if source incompatibilists have anything like the notion of an ineliminable explanation in mind when discussing the conditions of sourcehood.[22] But the promise alone of uniting leeway incompatibilism with source incompatibilism in a substantial way gives us another reason to take the explanatory solution seriously.[23]

[20] Thanks to Taylor Cyr for helping me see this.

[21] See Fischer (1994, ch. 7). Thank you to Carolina Sartorio for this point.

[22] See Pereboom (2001; 2014) again.

[23] See Kane (1996, pp. 73-75) as well as Sartorio (2015, pp. 264-266) on the prospects of uniting leeway and source incompatibilism.

References:

Campbell, Joseph K. (2007). “Free Will and the Necessity of the Past,” *Analysis* 67: 105–11.

Chisholm, Roderick (1966). “Freedom and Action,” in *Freedom and Determinism*, edited by Keith Lehrer. New York: Random House: 11-44.

Clarke, Randolph (1993). “Toward a Credible Agent-Causal Account of Free Will,” *Noûs* 27: 191–203.

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

Cutter, Brian (2017). “What is the Consequence Argument and Argument For?” *Analysis* 77: 278-287.

Fischer, John M. (1994). *The Metaphysics of Free Will: An Essay on Control.* Malden: Blackwell Publishing.

Fischer, John M. (2006). *My Way: Essays on Moral Responsibility*. New York: Oxford University Press.

Fischer, John M. (2011). “The Zygote Argument Remixed,” *Analysis* 71: 267-272.

Fischer, John M. & Ravizza, Mark (1998). *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press.

Frankfurt, Harry (1971). “Freedom of the Will and the Concept of a Person,” *Journal of Philosophy* 68: 5-20.

Franklin, Christopher (2011). "The Problem of Enhanced Control," *Australasian Journal of Philosophy* 89: 687-706.

Franklin, Christopher (2016). "If Anyone Should Be an Agent-Causalist, then Everyone Should Be an Agent-Causalist," *Mind* 125: 1101-1131.

Franklin, Christopher (2018). *A Minimal Libertarianism: Free Will and the Promise of Reduction*. Oxford: Oxford University Press.

Hobart, R.E. (1934). "Free Will as Involving Determination and Inconceivable Without It," *Mind* 43: 1–27.

Howard-Snyder, Daniel (2009). "The Puzzle of Thanksgiving and Praise," in *New Waves in Philosophy of Religion*, edited by Yujin Nagasawa and Erik J. Wielenberg. New York: Palgrave MacMillan: 125-149.

Kane, Robert (1996). *The Significance of Free Will*. Oxford: Oxford University Press.

Kearns, Stephen (2012). "Aborting the Zygote Argument," *Philosophical Studies* 160: 379-389.

Markosian, Ned (1999). "A Compatibilist Version of the Theory of Agent Causation," *Pacific Philosophical Quarterly* 80: 257–277.

Markosian, Ned (2012). "Agent Causation as the Solution to All the Compatibilist's Problems," *Philosophical Studies* 157: 383–398.

McKenna, Michael (2012). *Conversation and Responsibility*. New York: Oxford University Press.

Mele, Alfred (2006). *Free Will and Luck*. Oxford: Oxford University Press.

Nelkin, Dana K. (2011). *Making Sense of Freedom and Responsibility*. New York: Oxford University Press.

O'Connor, Timothy (2000). *Persons and Causes: The Metaphysics of Free Will*. New York: Oxford University Press.

O'Connor, Timothy (2005). "Freedom With a Human Face," *Midwest Studies in Philosophy* 29: 207–227.

O'Connor, Timothy (2009). "Agent-Causal Power," in *Dispositions and Causes*, edited by Toby Handfield. New York: Oxford University Press: 189–214.

Pereboom, Derk (2001). *Living without Free Will*. New York: Cambridge University Press.

Pereboom, Derk (2004). "Is Our Conception of Agent-Causation Coherent?" *Philosophical Topics* 31: 275–286.

Pereboom, Derk (2014). *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.

Reid, Thomas (1788/1969). *Essays on the Active Powers of Man*. Cambridge: MIT Press.

Sartorio, Carolina (2015). "The Problem of Determinism and Free Will is Not the Problem of Determinism and Free Will," in *Surrounding Free Will: Philosophy, Psychology, Neuroscience*, edited by Alfred Mele. Oxford: Oxford University Press: 255-273.

Sartorio, Carolina (2016). *Causation and Free Will*. New York: Oxford University Press.

Strawson, Galen (1994). “The Impossibility of Moral Responsibility,” *Philosophical Studies* 75: 5-24.

Todd, Patrick (2013). “Defending (a Modified Version of) the Zygote Argument,” *Philosophical Studies* 164: 189-203.

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

Watson, Gary (2004). *Agency and Answerability: Selected Essays*. New York: Oxford University Press.

Appendix: The Fixity of the Independent and Local-Miracle Compatibilism

Introduction

At the end of chapter 3, I asked whether classical compatibilists – those who think the freedom to do otherwise is compatible with determinism – should feel any pressure to accept the Principle of the Fixity of the Independent (FI). I claimed that at least *some* classical compatibilists should, but not all. The rationale was this: many of the arguments for FI start with the initial plausibility of the Principle of the Fixity of the Past (FP), the principle that no one is ever able to perform an action which requires the past to be different (roughly). Since some classical compatibilists seem to accept FP – so-called "local-miracle compatibilists" – it would seem as if such authors should feel pressure to accept FI. In contrast, classical compatibilists who explicitly reject FP – so-called "multiple-pasts compatibilists" – need not feel any pressure.[1]

In reality, the relationship between classical compatibilism, particularly local-miracle compatibilism, and FI is much more complicated. The goal of this appendix is to deal with those complications. In short, I'll be arguing that the most

[1] The phrases "local-miracle compatibilism" and "multiple-pasts" compatibilism are due to John Fischer (1994).

influential version of local-miracle compatibilism, the one taken from David Lewis (1981), is *not* susceptible whatsoever to the arguments for FI. However, Lewis's version faces some serious problems. And while Lewis's version can be amended so as to solve these problems, we end up with a version of local-miracle compatibilism that is susceptible to the arguments for FI. So while the *most popular* version of local-miracle compatibilism may be immune to the arguments for FI, the *most plausible* version is not.

It's a bit of a winding road to get there, so here's how we'll proceed. First, I'll lay out, in some detail, Lewis's version of local-miracle compatibilism and how it avoids the arguments for FI. Second, I'll highlight two problems for his version as well as promising solutions, thereby sketching a more plausible version of local-miracle compatibilism along the way. Finally, I'll show that this more plausible version, although capable of avoiding the problems of Lewis's version, falls prey to the arguments for FI.

Lewis's Local-Miracle Compatibilism and FI

At the heart of Lewis's (1981) version of local-miracle compatibilism is a seeming paradox. Suppose that, at time *t*, I refrain from throwing a rock and that my refraining was determined by the past and the laws. Lewis grants all three of the following claims: (i) I could have thrown the rock at *t*, (ii) if I had thrown the rock at *t*, a natural law would have been broken, and (iii) I cannot break any natural law.

Although these three claims might look inconsistent, Lewis dissolves the paradox by clarifying claims (ii) and (iii). Sure enough, he says, if I had thrown the rock, a law would have been broken, but my throwing the rock would not *itself* be, nor would it cause, a law-breaking event. Instead, some event *just prior* to my throwing the rock would have broken the laws. It may be difficult (if not impossible) to say what this law-breaking event would have been exactly – maybe it would have been a handful of particles briefly swerving in a different direction, or a few additional neurons firing, etc. But the important point is that it is the prior event, not my throwing the rock, that would have violated a law. He puts it like this:

> [My act] would not itself falsify any law – not if all the requisite law-breaking were over and done with beforehand. All that is true is that my act would be preceded by another event – the divergence miracle – that would falsify a law. (1981, p. 119)

So, says Lewis, (ii) is true because agents can perform actions which *require* preceding law-breaking events. But (iii) is also true because no agent can perform an action which *itself* would be, or cause, a law-breaking event. In Lewis's famous terminology, we ought to distinguish between the "Weak Thesis," which the local-miracle compatibilist should accept, and the "Strong Thesis," which she should reject:

> (Weak Thesis): I am able to do something such that, if I did it, a law would be broken.
>
> (Strong Thesis): I am able to break a law. (1981, p. 115)

By keeping these two theses distinct, the paradox above is resolved and the local-miracle compatibilist is not committed to any barbarous claims about what we are able to do.

That's the heart of Lewis's local-miracle compatibilism, but there are some natural follow-up questions. First, how can he guarantee that some prior event, and not the agent's action itself, will be the law-breaking event? If he can't guarantee this, then it seems the local-miracle compatibilist might have to accept both the Weak and the Strong Thesis. Second, how can he guarantee that the prior event will be *just* prior to the agent's action, not in the distant past instead? If he can't guarantee this, then the local-miracle compatibilist doesn't accept even the fixity of the *remote* past, meaning her view would qualify as a version of multiple-pasts compatibilism instead.

Lewis answers both questions in one fell swoop. In his other works, he provides a relatively detailed and independently motivated semantics for counterfactuals, one that seems to deliver exactly the results his local-miracle compatibilism needs.[2] His semantics starts with the claim that conditionals of the form "If *P* were true, *Q* would be true" hold just in case the closest world where *P*

[2] See Lewis (1973) and (1979).

is true, *Q* is also true (roughly).[3] Next, in order to tell us what makes one world "closer" than another, Lewis offers the following "similarity metric" for comparing worlds:

1. It is of the first importance to avoid big, widespread, diverse violations of law.
2. It is of the second importance to maximize the spatiotemporal region throughout which perfect match of particular fact prevails.
3. It is of the third importance to avoid even small, localized, simple violations of law.
4. It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly. (1979, p. 474)

Let's work through the example of my refraining from throwing the rock at *t*. Again, Lewis wants to say that if I had thrown the rock, then some law-breaking event *just prior* to my throwing the rock would have occurred. Now compare the following three worlds:

> *w_1*: Just before *t*, a few additional neurons fire in my brain, which causes me to throw the rock at *t*.

[3] More precisely, Lewis holds that it must be that for every world where *P* is true and *Q* is false, there is a closer world where *P* and *Q* are both true. We can ignore this complication in what follows.

> w_2: Everything is exactly the same all the way up until *t* – no additional neurons firing or anything like that – but I nevertheless throw the rock at *t*.
>
> w_3: Long before *t*, a handful of particles swerve in a slightly different direction, causing more particles to swerve, causing more particles… which eventually causes a few additional neurons to fire in my brain just before *t*, which causes me to throw the rock at *t*.

In order to get the desired result, it must be that w_1 is closer to the actual world than both w_2 and w_3. And sure enough, Lewis's similarity metric seems to imply exactly that. Right off the bat, w_1 fares better than w_2 because w_2 involves a *large* miracle. In w_2, my arm's movement is the law-breaking event, but my arm is made up of a great many particles, far more than those involved in a few additional neurons firing. So by rule (1), w_1 is closer than w_2. In regard to w_1 and w_3, these worlds seem to do equally well with respect to rule (1), as neither involves a large miracle. But w_1 does better with respect to rule (2): by having everything be *exactly* the same until just prior to *t*, w_1 has more *perfect* match with the actual world.

So using this similarity metric, Lewis seems to get the result that the law-breaking event would have occurred *just prior* to my throwing the rock. Now, admittedly, this is just one case, and we might be hesitant to generalize. But it's not implausible to do so. There would seem to be nothing special about me or my not throwing the rock. If so, then it looks like Lewis can secure the claim that, although our performing certain actions may *require* law-breaking events, those events are distinct from and just prior to such actions.

We are now in a position to see how Lewis's version of local-miracle compatibilism interacts with the arguments I gave for FI. Consider just the first argument presented in chapter 1 which ran, roughly, as follows: the best explanation for why the past is fixed, but not necessarily the future, is that the past is in no way explained by our current behavior whereas the future often is, and that any part of the world which isn't explained by our current behavior is fixed for us. But this latter claim, that any part of the world which isn't explained by our behavior is fixed for us, just is FI.

Lewis will find this wholly unconvincing. First, he will want to qualify the claim that the past is fixed for us. For Lewis, it is only the *remote* past which is fixed, not the *immediate* past. Second, the fixity of the remote past has nothing to do with its being explanatorily independent of our behavior. After all, the immediate past is explanatorily independent of our behavior, but it is *not* fixed for us. Instead, the fixity of the remote past has merely to do with the way we evaluate counterfactuals *à la* the similarity metric. The remote past is fixed, but not necessarily the remote future, because we want to maximize exact spatiotemporal match unless doing so requires a large miracle. And as Lewis argues,[4] no large

[4] See Lewis's (1979) response to Fine's (1975) objection involving future match. The central idea is that the future is, by definition, counterfactually dependent on the past in a more robust way than the past is on the future. Thank you to Michael Nelson here.

miracles are needed in order to maximize exact match with the remote past, whereas they often are needed in order to maximize exact match with the future.

So it seems as if Lewis's version of local-miracle compatibilism is immune to the arguments I presented for FI. I'll now present two shortcomings of his view as well as some promising solutions. I'll then argue that these solutions come at a cost for the local-miracle compatibilist: when taken together, they render her view susceptible to the arguments for FI.

The First Problem: Defining a "Law-Breaking" Event

The first problem, brought out in various ways by Terence Horgan (1985), Shane Oakley (2006) and Peter A. Graham (2008), centers around Lewis's resolution of the seeming paradox considered above. Recall that Lewis accepts the Weak Thesis but denies the Strong Thesis: he accepts that if I had thrown the rock, say, some law-breaking event would have occurred, although my throwing the rock would not itself be (or cause) a law-breaking event. The problem is that, using Lewis's own definition of a "law-breaking" event, *neither* the Weak Thesis nor the Strong Thesis is true: if I had thrown the rock, some non-actual event would have occurred, but it needn't be a *law-breaking* event, at least by Lewis's definition. And if we move away from Lewis's definition so as to ensure that the Weak Thesis comes out true, it is no longer clear that the local-miracle compatibilist needs to deny the Strong Thesis: it is not obvious she needs to avoid saying that my throwing the rock would itself be (or cause) a law-breaking event. Either way, and *contra* Lewis, the local-miracle compatibilist needn't accept the Weak Thesis while denying the Strong Thesis.

Start by considering how Lewis defines a "law-breaking" event:

> Let us say that an event would falsify a proposition iff, necessarily, if that event occurs then that proposition is false. (1981, p. 119)

A law-breaking event just is an event that falsifies the conjunction of propositions describing the laws: it is an event such that, necessarily, if that event occurs, then at least one of the laws does not obtain. With this definition in mind, it is indisputable that no one is able to perform an action that would itself be (or cause) a law-breaking event – that is, it is indisputable that the Strong Thesis is false. Peter van Inwagen (1983) provides an example to help illustrate this:

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

The bureaucrat’s demand is incoherent because a spaceship traveling faster than the speed of light is a “law-breaking” event in Lewis’s sense: necessarily, if that event occurs, then the laws of our world do not obtain.

The problem is that, while Lewis’s definition renders the Strong Thesis unacceptable, it seems to render the Weak Thesis dubious as well: it is not obvious that, if I were to throw the rock, a “law-breaking” event would have occurred. Suppose that if I had thrown the rock at *t*, a few additional neurons would have fired just prior to *t*. There is a big difference between a few additional neurons firing and a spaceship traveling faster than the speed of light: in contrast to the spaceship case, there are plenty of worlds with the same laws where a few additional neurons fire just prior to *t*. True, these worlds have different pasts, even remote pasts. But that’s neither here nor there with respect to this definition of a law-breaking event. If there is any world with the same laws where a few additional neurons fire just prior to *t*, then that event is not a law-breaking event according to Lewis’s definition. So although the local miracle-compatibilist should reject the Strong Thesis under Lewis’s definition, she needn’t accept the Weak Thesis.

This result might tempt us to try a different definition, one which would render the firing of a few additional neurons a law-breaking event, thereby justifying the Weak Thesis. Something like the following, inspired by van Inwagen, would do the trick:

> An event breaks law(s) *L* if, and only if, necessarily, if the event *and the entire actual past occur* then *L* doesn't obtain.[5]

Since every world where a few additional neurons fire just prior to *t* either have a different past or different laws, this definition would make it a law-breaking event. So this definition would commit the local-miracle compatibilist to the Weak Thesis. But as Terence Horgan (1985) and Christopher S. Hill (1992) point out, if the local-miracle compatibilist uses something like this definition, she should be open to accepting the Strong Thesis as well: she need not reject the claim that our actions *themselves* can break the laws. With this definition, to claim that our actions themselves cannot break the laws just is to claim that we can only perform actions that are consistent with the laws and the past. That's clearly begging the question against the classical compatibilist. At the very least, it's not supported by cases like van Inwagen's bureaucrat.[6] Setting aside Lewis's similarity metric for the moment, this means that the local-miracle compatibilist could, without any hesitation, insist that my throwing the rock would itself be the law-breaking event.

There are, of course, lots of other ways we might go about defining a "law-breaking" event, and the issue is quite delicate.[7] But the foregoing remarks suggest

[5] See van Inwagen (1983, p. 68; 2004).

[6] Although, see Fischer (forthcoming) for reasons to think we cannot perform actions which require law-breaking events, even in this sense.

[7] See Beebee (2003), Oakley (2006), Graham (2008), and Tognazzini (2016). Just as a quick example, consider Beebee's proposed definition in offering an argument against local-miracle compatibilism:

> Event *e* is a law-breaking event at world w_a, relative to world w_b, iff *e*, together with the circumstances under which it occurs at w_a, is incompatible with *L*, where *L* is the conjunction of all of w_b's laws. (2003, p. 266)

a more general strategy for the local-miracle compatibilist in answering the paradox we started with. When her view is charged with giving agents the ability to break the laws, she needn't follow Lewis and begin by distinguishing the Weak and Strong Theses. Instead, she should start by merely asking for a definition of a "law-breaking" event. Her opponent then has to "thread the needle" in offering a definition: go too far toward Lewis's definition, and the local-miracle compatibilist is not committed to the claim that we can perform actions that would themselves be (or even require) a law-breaking event; go too far toward van Inwagen's definition, and there isn't any obvious reason to think that we can't perform actions which would themselves be a law-breaking event. Without a definition that "threads the needle," the local-miracle compatibilist can diffuse the worry that her view imputes fantastic abilities to agents, and all without even mentioning the Weak and Strong Theses.[8]

This may seem like a small issue for the local-miracle compatibilist, but using this strategy over Lewis's brings with it three distinct advantages. First, it makes local-miracle compatibilism less vulnerable. As we saw, in order to fully justify the claim that the "law-breaking" event would have been some event distinct from my throwing the rock, Lewis had to appeal to his similarity metric. So if the

The trouble arises in specifying what the relevant "circumstances" are. They would have to be inclusive enough to render my throwing the rock a law-breaking event, but not so inclusive as to render the claim that no one can perform a law-breaking event question-begging. It's not immediately clear how to accomplish this.

[8] What if her opponent does produce a definition which threads the needle? Well then perhaps the local-miracle compatibilist ought to fall back on the distinction between the Weak and Strong Theses. But that's a big "if"! Moreover, I suspect the distinction won't help significantly, largely for the reasons Beebee (2003) gives. (See below.)

local-miracle compatibilist makes Lewis's strategy front and center, it looks as if the plausibility of her view is held hostage to particular details involving the semantics of counterfactuals. That should make the local-miracle compatibilist a little uneasy. Under the suggestion here though, she need only point out that however we define a "law-breaking" event, either her view will not countenance the ability to perform law-breaking events, or it will be question-begging (or at least unmotivated) to say we don't have such abilities. No counterfactual logic or similarity metric required.

The second advantage is that it gives the local-miracle compatibilist a straightforward response to an alleged counterexample from Helen Beebee (2003). Beebee argues that, if we shift our focus from *bodily* actions to *mental* ones, it's not so clear that Lewis can accept the Weak Thesis without also accepting the Strong Thesis. Here's how a modified version of the case goes. Suppose Lewis is right that, if I had thrown the rock at *t*, some "law-breaking" event ("law-breaking" in something like the second sense above) just prior to *t* would have occurred, like the firing of a few additional neurons. Also suppose, as seems immensely plausible, that if I had thrown the rock at *t*, I would have first *decided* to throw the rock – that worlds where I first *decide* to throw and then throw are closer than worlds where I decide to not throw but my arm moves against my will. The problem, according to Beebee, is that my *deciding* to throw the rock *just is* (or is at least "underwritten" by) a few additional neurons firing. That strongly suggests that the "local miracle" which would precede my throwing the rock just is an *action* of mine: that if I were to throw the rock, the preceding "law-breaking" event would be my *deciding* to throw the rock. If so, then it would seem as if Lewis cannot accept the Weak Thesis without also accepting the Strong Thesis.

There are a number of unappealing responses the local-miracle compatibilist could give. For instance, perhaps she could insist that even mental actions, such as my deciding to throw the rock, are too "large" to play the role of a local miracle. But this response seems to hinge on the wrong kinds of facts. Even if *our* mental actions still qualify as large miracles, that's only because we happen to be medium-sized physical objects. We can imagine smaller agents or even non-physical ones – think of Marvel's "Ant-Man" or a Cartesian soul – whose mental actions would arguably qualify as small miracles.[9] It would be odd if the local-miracle compatibilist escaped the objection simply because we happen to be a certain physical size. Or perhaps she could instead reply that, even though mental actions can be small miracles, we should posit even smaller ones preceding them so as to ensure a smooth transition. But as Beebee argues, this seems to lead to an infinite regress, each small miracle needing some earlier and smaller miracle.

Fortunately for the local-miracle compatibilist, the strategy above gives way to a more plausible response: she should instead simply admit that mental actions can play the role of a small "miracle," and so we can perform actions which themselves "break the laws" in *some sense*, but not in a *problematic sense*. Sure,

[9] Even non-agents would do the trick, like an amoeba swimming left rather than right. Even if this isn't an "action" properly speaking, we still don't want to say that mere amoeba can break laws.

none of us can perform actions which in themselves are inconsistent with the laws. But my deciding to throw the rock is nothing like that – there would seem to be plenty of worlds with the same laws where I decide to throw. More generally, the local-miracle compatibilist doesn't have to worry about accepting the Weak Thesis while denying the Strong Thesis, at least once the notion of a "law-breaking" event is defined. And if not, Beebee's challenge is far less compelling.

The third advantage of this strategy is related to this second advantage. By allowing some of our actions, such as mental actions, to serve as the requisite local "miracle," the local-miracle compatibilist can embrace the fixity of the *entire* past. As noted, one funny feature of Lewis's version is that he accepts the fixity of the *remote* past but not the *immediate* past. That might make Lewis's local-miracle compatibilism unattractive. It's bad enough, so the thought might go, that Lewis denies the fixity of the laws, but to deny the fixity of the past as well? That's too far. At least the multiple-pasts compatibilist can accept one of these intuitive principles – Lewis, strictly speaking, accepts neither![10]

The version of local-miracle compatibilism being suggested instead claims that I can now decide to throw the rock, say, but that my decision itself would be the local "miracle." That means the local-miracle compatibilist can admit that the closest world where I decide to throw the rock is one with the *exact* same past up until my decision. Admittedly, this strategy would only work for mental actions,

[10] Michael Huemer (2000, p. 541-544) seems to have something like this worry in mind. See Taylor Cyr (manuscript) for a more explicit statement of this point.

such as decisions, tryings, the forming of intentions, etc. But it wouldn't be too hard to come up with a version which made room for bodily actions as well. Most simply, she could allow for some bodily actions to count as small miracles. Or more plausibly, she could give a "tracing" account consonant with many incompatibilist theories. Roughly: I could have thrown the rock at *t* in virtue of the fact that there was some action I could have performed at some prior time, such as my deciding to throw the rock, that was consistent with the past relative to that prior time. Either strategy would allow the local-miracle compatibilist to embrace the fixity of the *entire* past, thereby making way for a "pure" version of the view, one that is sharply distinguished from any type of multiple-pasts compatibilism.

To sum up: Lewis, using his own definitions, was too quick to accept the Weak Thesis but deny the Strong Thesis. And rather than making the distinction between these two theses front and center, the local-miracle compatibilist ought to instead focus on defining the notion of a "law-breaking" event, as doing so brings with it three distinct advantages. First, her central response to the paradox above is no longer beholden to a similarity metric; second, she can offer a straightforward response to Beebee's alleged counterexample; third, she can accept the fixity of the entire past, ensuring her view is quite distinct from any version of multiple-pasts compatibilism. Let's now move on to the second problem for Lewis's version of local-miracle compatibilism: the similarity metric.

The Second Problem: The Similarity Metric

As we saw, Lewis uses his similarity metric to ward off two worries. First, he uses it to justify the claim that the divergent event would not be the agent's action itself, thereby ensuring that none of us can "break the laws" in a problematic sense. If the comments of the last section are on target, the local-miracle compatibilist need not worry about this. Second, Lewis uses it to justify the claim that the (remote) past would have been the same, thereby ensuring that his view is to be preferred to multiple-pasts compatibilism. The local-miracle compatibilist should still find this of some interest. For instance, the "pure" local-miracle compatibilism sketched above is committed to the claim that, if I had thrown (or decided to throw) the rock, the entire past would have been the same. If this counterfactual comes out false, as other similarity metrics imply,[11] then even this "pure" version is in trouble. Hence, even if Lewis's similarity metric is not necessary in answering the paradox we started with, it is still relevant in establishing local-miracle compatibilism over multiple-pasts compatibilism.

The problem is that Lewis's similarity metric has a serious flaw, one that centers around the fourth rule of his similarity metric:

4. It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly. (1979, p. 474).

[11] See Bennett (1984) and Dorr (2016).

The problem concerns the “little or no importance” clause – is approximate match worth something or not? It turns out that we get trouble either way.[12]

First, suppose we give *some* importance to approximate match. Paul Tichý (1976) presents a troublesome case: Jones always wears his hat when the weather is bad, but when the weather is good, he flips a fair coin to decide – heads, he wears it; tails, he doesn’t. Suppose that on Monday, the weather is bad and so Jones wears his hat. What would have happened if the weather had been nice? Intuitively, if the weather had been nice on Monday, Jones *might* have worn his hat or he might not have. But if approximate match has *some* weight, we don’t get this result: we are instead forced to say that he definitely *would have* worn his hat. To see this, compare the following two worlds:

> w_{hat}: The weather is nice on Monday, Jones flips his coin which comes up heads, and so Jones wears his hat.
>
> $w_{\text{no-hat:}}$ The weather is nice on Monday, Jones flips his coin which comes up tails, and so Jones doesn’t wear his hat.

In order to get the result that, if it had been nice on Monday, Jones might or might not have worn his hat, it must be that w_{hat} and $w_{\text{no-hat}}$ are equally close to the actual world. But if we assign some weight to approximate match, then we don’t get that result. Since Jones wears his hat on Monday in the actual world, any world in which

[12] The presentation of this problem is taken almost completely from Wasserman (2018, ch. 4), although he does not use the problem to support FI.

Jones wears his hat on Monday has some approximate similarity to the actual world. And because w_{hat} and $w_{\text{no-hat}}$ are tied with respect to rules (1), (2), and (3), this means w_{hat} is closer to the actual world. That's the wrong result.

Suppose instead that we assign *no* weight to approximate match. This avoids Tichý's hat case, with w_{hat} and $w_{\text{no-hat}}$ now being equally close, but we face another troublesome case given by Sidney Morgenbesser through Michael Slote (1978, n. 33): suppose that your friend flips a coin, and while it is in the air, he bets you a large sum of money that it will land tails. You are a bit risk-averse and so refuse the bet. When the coin hits the ground, you immediately regret your refusal, the coin landing heads. Intuitively, it seems that, if you had taken the bet, you would have won – that is why you experience regret, after all. But if we assign *no* weight to approximate match, Lewis's similarity metric doesn't deliver this counterfactual. Compare two worlds:

> $w_{\text{bet+heads}}$: You take the bet and the coin lands heads, winning you the bet.
>
> $w_{\text{bet+tails}}$: You take the bet and the coin lands tails, losing you the bet.

In order to get the result that, if you had taken the bet, you would have won, it must be that $w_{\text{bet+heads}}$ is closer than $w_{\text{bet+tails}}$. Now both worlds match *perfectly* up until you decide to bet, but after that, $w_{\text{bet+heads}}$ only has better *approximate* match. Since these worlds are tied with respect to rules (1), (2), and (3), if approximate match has *no* weight, then there is no reason to think $w_{\text{bet+heads}}$ is closer to the actual world. Again, we get the wrong result.

Fortunately for the local-miracle compatibilist, there is a relatively simple and elegant amendment to Lewis's similarity metric that avoids the dilemma, one first proposed by Jonathan Schaffer (2004). It starts by noticing a *causal* difference between Tichý's hat case and Morgenbesser's betting case: whereas the bad weather *causes* Jones to wear the hat, your refusing the bet does *not cause* the coin to land heads. Schaffer's idea is that we don't want to maximize match (exact or approximate) with respect to *every* region, but only regions *that are causally independent of the antecedent*. We don't care about worlds where the weather is nice but Jones is wearing his hat because his wearing the hat is causally *dependent* on the weather. But we do care about worlds where you take the bet and the coin lands heads because the coin landing heads is causally *independent* of your taking the bet. More formally, we can amend Lewis's similarity metric as follows:

1*. It is of first importance to avoid big miracles.

2*. It is of second importance to maximize the region of perfect match, *from those regions causally independent of whether or not the antecedent obtains*.

3*. It is of third importance to avoid small miracles.

4*. It is of fourth importance to maximize the spatiotemporal region of approximate match, *from those regions causally independent of whether or not the antecedent obtains*. (Schaffer 2004, p. 305)

Schaffer's amendment is an elegant solution to the problem posed by Tichý and Morgenbesser, one that gives the local-miracle compatibilist a principled reason for claiming that the remote past is fixed, *contra* the multiple-pasts compatibilist. And if combined with the "pure" local-miracle compatibilism above, she can justify the claim that the *entire* past is fixed in evaluating what we can and can't do. Consider my deciding to throw the rock again. In evaluating what would have happened had I decided to throw the rock, rules (1*) and (2*) tell us to maximize exact match with respect to regions that are causally independent of my decision unless doing so requires a large miracle. The past is clearly causally independent of my decision. And, as I argued earlier, my decision to throw the rock need only be a small miracle. Hence, worlds where my decision is the first divergent event will always be closer than worlds where the first divergent event is in the past, even the immediate past. If so, then Schaffer's similarity metric implies that, if I had decided to throw the rock, the *entire* past up until my decision would have been the same.

So by invoking the notion of causation, the local-miracle compatibilist can save Lewis's similarity metric from some of its most pressing problems. This is not to say the metric faces *no* problems.[13] Moreover, Lewis certainly wouldn't like the revisions.[14] But the metric is at least much more plausible. Additionally, it gives the "pure" local-miracle compatibilist a deep reason for thinking the *entire* past is

[13] See Dorr (2016) again.

[14] As Schaffer (2004) points out, this revised metric would make Lewis's counterfactual analysis of causation circular.

fixed, thereby separating her view sharply from the multiple-pasts compatibilist. All of this taken together seems to make for a much more compelling version of local-miracle compatibilism. However, as I'll now finally argue, this much more compelling version faces one big problem: it is susceptible to the arguments for FI.

Back to FI

Recall how the argument for FI went: the best explanation for why the past is fixed, but not necessarily the future, is that the past is in no way explained by our current behavior, and any part of the world which is in no way explained by our current behavior is fixed for us. But this latter claim just is FI.

As we saw, Lewis will contest every step of the argument here. According to Lewis, it is only the *remote* past, not the *immediate* past, that is fixed for us. Moreover, the fixity of the (remote) past has nothing to do with FI. Instead, the fixity of the (remote) past just has to do with Lewis's similarity metric and its insistence that we maximize spatiotemporal match over *all* regions.

It is a much different story under the local-miracle compatibilism we've arrived at. First, the "pure" local-miracle compatibilist accepts, or is at least open to, the fixity of the *entire* past: she accepts, or is at least open to, the claim that there is no action we can perform such that, if we were to perform it, any part of the past would be different. So the first step of the argument goes through. Second, by using the similarity metric Schaffer proposes, her explanation for the fixity of the past, but not the future, looks extremely similar to FI: the entire past is fixed, but not the future, because the past is not caused by our current behavior, and we ought to maximize match with respect to those regions that are causally independent of our current behavior. Granted, this explanation only admits that those parts of the world that are *causally* independent of our behavior are fixed, not necessarily those that are *explanatorily* independent. But as I argued in the first chapter, it's a short step from causation to explanation, one that the local-miracle compatibilist doesn't have any special resources to avoid.[15] At the very least, the local-miracle compatibilist will now stay "onboard" with the argument considerably longer than Lewis.

The local-miracle compatibilist, as sketched here, seems to be in a precarious position. She admits that the entire past is (or may very well be) fixed in evaluating what we can currently do, and that this fixity stems from the past's being causally independent of our current behavior. However, she *denies* that the laws are fixed. But why the difference? After all, it would seem as if the laws are also causally independent of anything we do (setting aside Humean views of the laws).[16] True, the laws are not *counterfactually* independent of our behavior, at least not if Schaffer's similarity metric is right. But that's neither here nor there. It still seems as if the laws don't obtain *in virtue of* anything we do. If the past is fixed

[15] Recall that the step from causation to explanation involved certain asymmetries around so-called "soft facts." I don't see how the local-miracle compatibilist has anything special to say here.

[16] See chapter 3 for discussion of Humeanism and its relation to FI. Perhaps it's no coincidence that Lewis (1994) endorsed a Humean view of the laws.

because it is causally independent of our current behavior, why wouldn't the laws be fixed as well? The local-miracle compatibilist owes us a story here.

It seems to me the most the local-miracle compatibilist can say is that, since freedom is compatible with determinism, *something* has to give if we were to do otherwise in a deterministic setting; Schaffer's similarity metric tells us it's the laws rather than the past. So long as this doesn't imply any barbarous claims about our abilities, that's good enough. But this response, apart from being somewhat question-begging, highlights what's so attractive about incompatibilism as driven by FI: the Principle of the Fixity of the Independent *accounts* for more familiar principles like the Principle of the Fixity of the Past and the Principle of the Fixity of the Laws. This gives incompatibilism more explanatory coherence, making it a more satisfactory view. Perhaps local-miracle compatibilism, as sketched here, doesn't say anything absurd, but it seems to lack the theoretical unity that incompatibilism as driven by FI has. All else being equal, that makes incompatibilism more appealing.

Concluding Remarks

As you can see, the relationship between local-miracle compatibilism and FI is quite complex, and a good amount depends on the details. In this chapter, I've tried to sketch a more plausible version of local-miracle compatibilism, one that is in the spirit of Lewis's but avoids some of its pitfalls. However, this more plausible version leads us right back into the arms of FI. Thus, it seems there is some pressure for the local-miracle compatibilist to abandon her classical compatibilism and accept that the truth of determinism would undermine our being free to do otherwise.

References:

Beebee, Helen (2003). "Local Miracle Compatibilism," *Noûs* 37: 258-277.

Bennett, Jonathan (1984). "Counterfactuals and Temporal Direction," *Philosophical Review* 93: 57-91.

Cyr, Taylor. "A Dilemma for Local Miracle Compatibilism," manuscript.

Dorr, Cian (2016). "Against Counterfactual Miracles," *Philosophical Review* 125: 241-286.

Fine, Kit (1975). "Critical Notice of Lewis's *Counterfactuals*," *Mind* 84: 451-458.

Fischer, John M. (1994). *The Metaphysics of Free Will: An Essay on Control*. Malden: Blackwell Publishing.

Fischer, John M. (forthcoming). "Local-Miracle Compatibilism: A Critique," in *Free Will: Historical and Analytical Perspectives*, edited by Jörg Noller and Marco Hauser. London: Palgrave/Macmillan Press.

Graham, Peter A. (2008). "A Defense of Local Miracle Compatibilism," *Philosophical Studies* 140: 65-82.

Hill, Christopher S. (1992). "Van Inwagen on the Consequence Argument," *Analysis* 52: 49-55.

Horgan, Terence (1985). "Compatibilism and the Consequence Argument," *Philosophical Studies* 47: 339-356.

Huemer, Michael (2000). "Van Inwagen's Consequence Argument," *Philosophical Review* 109: 524-544.

Lewis, David (1973). *Counterfactuals*. Malden, Massachusetts: Blackwell Publishing.

Lewis, David (1979). “Counterfactual Dependence and Time’s Arrow,” *Noûs* 13: 455-476.

Lewis, David (1981). “Are We Free to Break the Laws?” *Theoria* 47: 113-121.

Lewis, David (1994). “Humean Supervenience Debugged,” *Mind* 103: 473-490.

Oakley, Shane (2006). “Defending Lewis’s Local Miracle Compatibilism,” *Philosophical Studies* 130: 337-349.

Schaffer, Jonathan (2004). “Counterfactuals, Causal Independence, and Conceptual Circularity,” *Analysis* 64: 299-309.

Slote, Michael (1978). “Time in Counterfactuals,” *Philosophical Review* 87: 3-27.

Tichý, Paul (1976). “A Counterexample to the Stalnaker-Lewis Analysis of Counterfactuals,” *Philosophical Studies* 29: 271-273.

Tognazzini, Neal A. (2016). “Free Will and Miracles,” *Thought* 5: 236-238.

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

van Inwagen, Peter (2004). “Freedom to Break the Laws,” *Midwest Studies in Philosophy* 28: 334-350.

Wasserman, Ryan (2018). *Paradoxes of Time Travel*. Oxford: Oxford University Press.