Title
Ability and Responsibility: Essays on Behalf of Leeway Incompatibilism

Permalink
https://escholarship.org/uc/item/6nr33781

Author
Swenson, Philip

Publication Date
2014

Peer reviewed|Thesis/dissertation
Ability and Responsibility: Essays on Behalf of Leeway Incompatibilism

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

Doctor of Philosophy
in
Philosophy
by
Philip Swenson

August 2014

Dissertation Committee:
Dr. John Martin Fischer, Chairperson
Dr. John Perry
Dr. Michael Nelson
The Dissertation of Philip Swenson is approved:

                                       
                                       
                                       
                                       Committee Chairperson

University of California, Riverside
Acknowledgements

So many people have contributed to my becoming a philosopher. Collectively they are owed an unrepayable debt.

It begins with my parents. They cultivated an environment in which asking questions and seeking to understand why things were the way they were was accepted and encouraged. I had some natural philosophical proclivities. I remember worrying about whether I could know that I was awake rather than dreaming before age 8. But without my parents encouraging these tendencies, I doubt I would have ended up in philosophy.

My dad and I had a great many discussions that cultivated my love for ideas and arguments. He knew a bit of philosophy. I remember him telling me about the thought of Descartes and Alvin Plantinga. I got enough of a grip on philosophy from these discussions to realize it was something I very much wanted to learn more about.

My mom took on the massive burden of homeschooling my siblings and I. She devoted countless hours to my education but she also scheduled plenty of time to read on our own. I attribute my love of reading (a useful quality in philosophy) to her reading based approach to education. Both of my parents also strongly encouraged me in intellectual pursuits. They gave me the confidence, interest and education needed to go into philosophy.

My siblings also contributed significantly to my love of ideas. Intellectual debate around the family dinner table was (and still is) a common occurrence. My sister Alissa,
being closest in age, was a particularly good source of spirited discussion of all sorts of topics.

I did not major in philosophy as an undergraduate, so I entered graduate school at the University of Missouri excited about philosophy, but with extremely limited knowledge of the field. Luckily for me I found, both at Missouri and later at the University of California Riverside, and incredible (and incredibly talented!) group of fellow graduate students from whom to learn. Members of my incoming class at Missouri included a bunch of guys who created a great philosophical community. I learned tons of philosophy hanging out in the philosophy lounge talking with Kevin McCain, Andrew Moon, Patrick Todd, Kenny Boyce, Garrett Pendergraft, Robert Perez, Brandon Schmidly and others. All of them knew more philosophy than I did and hanging out with them put me in an incredibly friendly and intellectually rich environment. I learned at least as much philosophy from them as I did from seminars! Students from other classes from whom I also learned a lot from include Eric Roark, Alan Tomhave and Michael Hartsock, Jon Vertanen and Chris Gadsden. Justin McBrayer was my assigned mentor and went above and beyond in disabusing me of bad ideas and showing me the ropes of philosophy. Andrew Moon also taught me a lot. And his friendship and encouragement (both then and now) has been invaluable. Professors I learned a lot from at Missouri include Peter Markie, Peter Vallentyne, Paul Weirich, Robert Johnson, and especially Brian Kierland (with whom I had numerous great discussions outside of class).

When I arrived at UC Riverside I was lucky enough to find another incredible group of grad students working on free will and moral responsibility: Neal Tognazzini,
Chris Franklin, Ben Mitchell-Yellin, Justin Coates, Patrick Todd, Garrett Pendergraft, Joshua Hollowell and others. This was another great community to learn and grow in! Jason Grey, Heinrik Hellwig and Jonah Nagashima were also great sources of philosophical discussion during my Riverside years. Justin Coates and I had innumerable great philosophical discussions. Though we most often disagreed, I greatly enjoyed and benefited immensely from the dialogue and friendship. Patrick Todd was at Missouri as well as UCR. I was lucky to have him as a colleague for much of my grad school career. I have greatly appreciated his friendship. We have similar views on many issues and it has been a privilege to get to think through so many philosophical questions with him.

Professors I learned a lot from at UCR include Paul Hoffman, Coleen Macnamara and John Perry. Michael Nelson was a particularly great source insight (in and out of seminar).

Since I arrived at Riverside, John Martin Fischer has been as good of an advisor and mentor as one could ask for. In addition to being a first class philosopher, John is highly devoted to his students. He is incredibly generous with his time: both in giving quick written feedback on work and in meeting in person regularly. He has also been incredibly helpful and supportive of me in my pursuing of professional opportunities (such as fellowships and jobs) and has been a wonderful source of encouragement throughout my time in grad school. This dissertation would not have been possible without him.

And lastly, my wife Marie has been an essential part of this process. I am tremendously thankful for her willingness to move across the country to Riverside. She
has been with me through the ups and downs of grad school. She has displayed great love and patience throughout this journey and has been a wonderful source of comfort and support. I am incredibly grateful!

****

Chapter 3, in part, reprints material from my paper “A Challenge for Frankfurt-Style Compatibilists” forthcoming in *Philosophical Studies*. I thank the publisher for allowing me to use this material.
Dedication

For Marie
ABSTRACT OF THE DISSERTATION

Ability and Responsibility: Essays on Behalf of Leeway Incompatibilism

by

Philip Swenson

Doctor of Philosophy, Graduate Program in Philosophy
University of California, Riverside, August 2014
Dr. John Martin Fischer, Chairperson

Leeway incompatibilism is the view that (a) no one is morally responsible for what they do unless they could have done (or had the ability to do) something other than what they did and (b) the ability to do otherwise is incompatible with the obtaining of causal determinism. In this dissertation I attempt to defend the plausibility of leeway incompatibilism by investigating a variety of philosophical issues. I respond to arguments against leeway incompatibilism and develop accounts of certain phenomena (such as the nature of ability and the content of the blaming emotions or attitudes) that fit nicely with the leeway incompatibilist picture.

I have 5 chapters in the dissertation. The first chapter is introductory. The second and third chapters deal with the attack on the Principle of Alternative Possibilities (PAP) derived from the Frankfurt-style cases. Chapter two deals with ‘indirect’ attempts to use Frankfurt-style cases (FSCs) to undermine PAP. Proponents of Indirect Frankfurt-Style Arguments can (but need not) grant that there is no FSC in which an agent is both morally responsible and lacking alternative possibilities. They maintain that even if this is
so, the FSCs provide us with an anti PAP insight. I respond to several of these arguments.

Chapter three develops an original critique of the Frankfurt-style case based attack on PAP. This critique involves a substantial discussion of responsibility for omissions. In Chapter four I respond to an argument that aims to show that agents do not need to be in control of their behavior in order to be appropriately blamed for what they do. This involves developing an original account of the content of blaming attitudes (such as guilt and resentment).

In chapter five I develop an account of ability which yields the result that the ability to do otherwise is incompatible with causal determinism. This account of ability makes substantial use of the notion of ‘explanatory dependence’ which has been developed in recent philosophical literature. An interesting upshot of my account of ability is that foreknowledge of what action a person will perform is compatible with the person having the ability to do otherwise.
# Table of Contents

Chapter 1: Introduction………………………………………………………………..1

Chapter 2: Are Indirect Frankfurt-Style Arguments Persuasive?………………….7

Chapter 3: Omissions and the Frankfurt Cases: A New Challenge………………31

Chapter 4: Control, Blame and the Reactive attitudes…………………………..63

Chapter 5: Ability, Forknowledge and Explanatory Dependence………………..88
Chapter 1: Introduction

Leeway incompatibilism is the view that (a) no one is morally responsible for what they do unless they could have done (or had the ability to do) otherwise and (b) the ability to do otherwise is incompatible with the obtaining of causal determinism. Both of the claims that jointly make up leeway incompatibilism have been thought highly plausible by many philosophers as well as many outside of philosophy. I myself find leeway incompatibilism to be a natural way of looking at things. In fact, the idea that (b) might be false did not so much as occur to me until I entered graduate school. I take it that there are elements of common sense that strongly support leeway incompatibilism (although common sense may not speak with one voice on these issues.) Despite the naturalness of the picture, however, it faces a variety of significant philosophical difficulties. This dissertation aims to defend and increase the plausibility of leeway incompatibilism by investigating a variety of philosophical issues. This will involve responding to arguments against aspects of leeway incompatibilism and developing accounts of certain phenomena (such as the nature of ability and the content of the blaming emotions or attitudes) that fit nicely with the leeway incompatibilist picture.

Leeway incompatibilism faces several distinct challenges. My hope is that each chapter of this dissertation will go some way towards answering one or another of the challenges to leeway incompatibilism. Chapters 2 and 3 will focus on what I will call the Frankfurtian challenge.
This challenge originates from an extremely influential paper by Harry Frankfurt (1969). This paper introduced Frankfurt-style cases (FSCs). FSCs are examples intended to show that a person can be morally responsible for an action that they could not avoid: thus undermining one of the two central claims of leeway incompatibilism. Here is a standard FSC:

Black wishes Jones to cast his vote for presidential candidate A. In order to ensure that Jones does this, he implants a chip in Jones’s brain which allows him to control Jones’s behavior in the voting booth. (Jones has no idea about any of this.) Black prefers that Jones vote for candidate A on his own. But if Jones starts to become inclined to vote for anyone other than A, Black will immediately use his chip to cause Jones to vote for candidate A instead. As it turns out, though, Jones votes for candidate A on his own and Black never exerts any causal influence on Jones’s behavior.

Here it is intuitive that Jones is morally responsible for voting for candidate A despite the fact that it appears that he could not do otherwise. The early version of the Frankfurtian Challenge claims that FSCs are counterexamples to the claim that no one is morally responsible for what they do unless they could have done otherwise. Later (indirect) versions of the challenge claim that we can glean an anti-leeway incompatibilist lesson from the FSCs even if the FSCs are not counterexamples to PAP. The idea here is that reflecting on the FSCs help show us that possessing the ability to do otherwise is irrelevant to whether an agent is morally responsible. And the FSCs show us this even if
FSCs are not counterexamples to PAP. In Chapter 2 I will attempt to undermine several ways of pressing this indirect version of the Frankfurtian Challenge.

In Chapter 3 I will offer a second line of response to the Frankfurtian Challenge. The FSCs get there traction from the fact that Black does not make anything happen in the actual sequence of events. He plays no role in causing Jones to act as he does. The intuitive principle behind the FSCs seems to be something like: factors that are causally irrelevant are irrelevant to moral responsibility. (This principle is formulated more rigorously in chapter 3.) Given this principle and Black’s causal irrelevance, it seems that Black does not impact Jones’s responsibility. However, as I argue in chapter 3, there are clear counterexamples to the principle that factors that are causally irrelevant are irrelevant to moral responsibility. Given this, I argue that (despite initial appearances) we have good reason to deny that the agents in FSCs are morally responsible. I examine cases in which it is clear that an agent is not morally responsible for his behavior (or lack thereof) and argue that our judgment concerning responsibility in these cases ought to be the same as our judgment in the Frankfurt-style cases.

So chapters 2 and 3 develop responses to the Frankfurtian challenge. In Chapter 4 I will respond to one version of what I’ll call the Strawsonian challenge to leeway incompatibilism. PF Strawson’s extremely influential article, ‘Freedom and Resentment’(1962) drew our attention to the importance the reactive attitudes (e.g. attitudes such as resentment, indignation and guilt) for our theorizing about moral responsibility. Strawson thought that proper reflection on the nature of the reactive attitudes and their role in our moral responsibility practices reveals that the reactive
attitudes do not require justification of the sort suggested by leeway incompatiblists (i.e. the agent had the ability to do otherwise). If we take the plausible view that being morally responsible is a matter of being an appropriate target of the reactive attitudes, then Strawson’s view conflicts with leeway incompatibilism.

So by a *Strawsonian challenge* I mean any challenge that claims that reflection on the nature of the reactive attitudes and their role in our moral responsibility practices should lead us to the conclusion that leeway incompatibilism is mistaken. There are a variety of ways one might advance a Strawsonian challenge. In chapter 4 I will consider a Strawsonian challenge that has been pressed by Pamela Hieronymi (2004). Hieronymi raises a significant challenge to those (e.g. leeway incompatibilists) who think that the reactive attitudes (such as resentment and guilt) are only appropriate if the agent towards whom they are directed had robust control over her behavior.

The challenge goes something like this: Consider some (non-blaming) negative attitudes we can take towards others. Distrust and fear are good examples. Taking these attitudes towards an agent is often harmful to them but can nonetheless be justified even if the agent did not have robust control over her behavior. Why then are the reactive attitudes different? Why do they require robust control? In chapter 4 I attempt to answer these questions. I argue that the attitudes of resentment, guilt and indignation all essentially have content that entails that the agent towards whom they are directed had control over her behavior. That is why they are inappropriate in the absence of control.

Chapter 5 of the dissertation will take up what I will call *the challenge from freedom/foreknowledge compatibilism*. Unlike the claims underlying the Frankfurttian and
Strawsonian challenges, it might seem odd to think that an endorsement of the compatibility of freedom and foreknowledge would create a problem for leeway incompatibilists. Freedom/foreknowledge compatibilism says that foreknowledge and the ability to do otherwise are compatible. Leeway incompatibilism says that the ability to do otherwise and determinism are incompatible. What’s the problem?

The challenge arises because one of the primary motivations for accepting leeway incompatibilism also seems to motivate rejecting freedom /foreknowledge compatibilism. A central motivation for leeway incompatibilism is the idea that if facts beyond my control entail that I will perform a certain act, then I cannot act otherwise. But this claim appears to be in tension with freedom /foreknowledge compatibilism. Suppose God believed 1000 years ago that Jones will do X tomorrow, given God’s essential omniscience, it follows that Jones will do X. So it appears that leeway incompatibilists must either say that what God believed in the past was up to Jones or deny that Jones can act otherwise.

One reaction to the forgoing would be to say, ‘no great challenge here, leeway incompatibilists should also be freedom /foreknowledge incompatibilists.’ However, there are at least two costs to this response. First, freedom /foreknowledge incompatibilism is strange in a way that leeway incompatibilism is not. Incompatibilism about freedom and determinism is plausible in part because it is easy to see that causal determinism is the kind of thing that could generate limits on our abilities. If e is causally

---

1 This sort of claim forms the basis for the popular ‘consequence argument’ for incompatibilism.
determined to occur then there are causal forces at work in the world which will prevent me from acting in ways inconsistent with the occurrence of e.

But in the case of foreknowledge things are a bit more mysterious. Suppose that God knows that e will occur. How would God’s knowing this prevent me from acting in ways inconsistent with the occurrence of e? It would appear that freedom/foreknowledge incompatibilists would need to either claim that God’s foreknowledge is always backed by causal forces or posit a mysterious metaphysical mechanism to explain why foreknowledge removes the ability to do otherwise.

The second cost of saying that leeway incompatibilists should also be freedom/foreknowledge incompatibilists is just that it forces leeway incompatibilists to take on a further commitment. Evidence against freedom/foreknowledge incompatibilism is now evidence against leeway incompatibilism. Furthermore, those attracted to freedom/foreknowledge compatibilism will be disinclined to accept leeway incompatibilism.

So it would be advantageous if leeway incompatibilists could avoid being committed to freedom/foreknowledge incompatibilism. In chapter 4 I will attempt to dissolve the tension between leeway incompatibilism and freedom/foreknowledge compatibilism by developing an account of ability that supports them both.
Chapter 2: Are Indirect Frankfurt-Style Arguments Persuasive?

In 1969 Harry Frankfurt presented an extremely influential challenge to the Principle of Alternative Possibilities (PAP).\(^2\) This principle tells us that an agent is morally responsible for an action only if he could have done otherwise. PAP strikes most of us as initially plausible. Frankfurt, however, called PAP into question by presenting a case in which an agent appeared to morally responsible for an action despite the fact that he could not have acted otherwise. John Martin Fischer provides us with the following (politically modernized) version of Frankfurt’s case:

Because Black dares to hope that the Democrats finally have a good chance of winning the White House, the benevolent but elderly neurosurgeon, Black, has come out of retirement to participate in yet another philosophical example… He has secretly inserted a chip in Jones’s brain which enables Black to monitor and control Jones’s activities. Black can exercise this control through a sophisticated computer which he has programmed so that, among other things, it monitors Jones’s voting behavior. If Jones were to show any inclination to vote for McCain (or, let us say, anyone other than Obama), then the computer, through the chip in Jones’s brain, would intervene to assure that he actually decides to vote for Obama and does so vote. But if Jones decides on his own to vote for Obama (as Black, the old progressive would prefer), the computer does nothing but continue to monitor—without affecting—the goings-on in Jones’s head. Now suppose that Jones decides to vote for Obama on his own, just as he would have if Black had not inserted the chip in his head. It seems, upon first thinking about this case, that Jones can be held morally responsible for this choice and act of voting for Obama, although he could not have chosen otherwise and he could not have done otherwise.\(^3\)

Cases like this (commonly referred to as Frankfurt-style cases or FSCs) have generated an impressive literature on the question of whether they are in fact


counterexamples to PAP. One particular issue that has arisen concerns how we should fill in the details of the FSCs. Should we describe them in such a way that causal determinism obtains in the actual sequence? That is, should we hold that it is causally determined that Jones will vote for Obama even without bringing into play the presence of Black’s counterfactually intervening computer? Or should we posit causal indeterminism at a crucial point in the sequence leading to Jones’s vote? Proponents of the Dilemma Defense insist that the devil is in this particular detail and that neither the deterministic nor indeterministic option can provide us with a genuine counterexample to PAP.  

In short, the Dilemma Defense holds that in indeterministic FSCs the presence of the counterfactual intervener cannot remove all alternative possibilities. Deterministic FSCs, on the other hand, beg the question against those who hold that moral responsibility is incompatible with the truth of causal determinism. This is because, once we stipulate that causal determinism obtains in the actual sequence, incompatibilists will no longer find it plausible that Jones is morally responsible for voting for Obama.

Defenders of Frankfurt style arguments against PAP have offered a variety of replies to this dilemma. Some have attempted to craft more sophisticated indeterministic FSCs in which the agent lacks alternative possibilities that are robust enough to be

---

relevant to moral responsibility.\textsuperscript{5} Others have attempted to show that we can generate FSC-based arguments against PAP even if the FSCs do not provide us with a clear counterexample to PAP. Arguments of the latter sort are Indirect Frankfurt-Style Arguments. Proponents of Indirect Frankfurt-Style Arguments can (but need not) grant the success of the dilemma defense in showing that there is no FSC in which Jones is both morally responsible and lacking alternative possibilities. They maintain that even if this is so, the FSCs provide us with an anti PAP insight.

My purpose is to examine the prospects of Indirect Frankfurt-Style Arguments. My conclusion will be that, as of yet, the best attempts at generating a problem for PAP by appealing to Indirect Frankfurt-Style Arguments are less than persuasive. I will examine two prominent strategies for generating Indirect Frankfurt-Style Arguments: strategies that appeal to explicitly deterministic FSCs and strategies that grant that the agent possesses robust alternative possibilities in FSCs. I begin with the first strategy.

II

Since Fischer has led the way in developing anti-PAP arguments that make use of deterministic FSCs and since I take it that his general account offers the best hope of success, I will focus primarily on his account in what follows.\textsuperscript{7} I do hope to show that


\textsuperscript{6} But note that this conclusion does not entail that PAP is false. As we will see below, it might be the case that alternative possibilities are in and of themselves irrelevant to responsibility, but that their presence is nonetheless guaranteed by something else that is a necessary condition for an agent’s being morally responsible. In that case PAP would turn out to be correct.

worries I raise for his account are worries that must be overcome by anyone who wishes to undermine PAP by appealing to deterministic FSCs.

On Fischer’s view we can, without begging the question against the incompatibilist by assuming that Jones is responsible, utilize deterministic FSCs to demonstrate that the mere lack of alternative possibilities is not relevant to one’s moral responsibility. Fischer’s approach is to appeal to the seeming irrelevance of the presence of Black (and his counterfactually intervening device) to Jones’s responsibility. Since Black’s device is irrelevant to Jones’s responsibility, it appears that if Black’s device rules out Jones’s ability to do otherwise, then the mere fact that something rules out the ability to do otherwise is not relevant to responsibility. Fischer has argued in the following manner:

1. Assume that causal determinism obtains and that the Frankfurt case of Jones and Black unfolds as above.
2. Black’s presence (and device) and dispositions in themselves and apart from the assumption of causal determinism rule out Jones’s access to alternative possibilities. [Causal determinism in itself and apart from Black also in itself rules out alternative possibilities, but this will not be pertinent here.]
3. Black’s presence (and device) and dispositions in themselves are irrelevant to moral responsibility.

Thus, from 2 and 3:

4. Lack of alternative possibilities is in itself irrelevant to moral responsibility.

Thus:

5. If causal determinism rules out moral responsibility, it is not in virtue of eliminating access to alternative possibilities.\(^8\)

---

\(^8\) This formulation of the argument is taken from Fischer 2010 p. 324-325.
(4) and (5) are both quite significant conclusions. If this argument is sound, deterministic FSCs do present a serious challenge to the relevance of alternative possibilities to moral responsibility. I agree with Fischer that it is intuitive that (3) is true. Black’s presence is irrelevant to Jones’s responsibility. However, this argument, as currently formulated, faces a serious objection.

The problem is that (2) does not appear to be true. Recall that if we do not assume that causal determinism holds then it will not be true that the presence of Black’s device eliminates all of Jones’s alternative possibilities. Thus it appears that Black and his device do not eliminate Jones’s access alternative possibilities “in themselves and apart from the assumption of causal determinism.” Stewart Goetz presses this objection as follows:

[The deterministic FSC] creates the appearance that it is Black’s device, which is in the alternative sequence of events, that makes it the case that Jones is not free to choose otherwise. This appearance is illusory because without the obtaining of causal determinism in the actual sequence of events, the device cannot prevent Jones from making an alternative choice, and with causal determinism in the actual sequence of events it is not the device that prevents Jones from making an alternative choice. In short, if Jones is not free to choose otherwise, it is because of the occurrence of causal determinism in the actual sequence of events and not because of Black’s device in the alternative sequence. Goetz also suggests that Fischer’s argument begs the question against the incompatibilist. While I am not convinced of this claim (none of the premises in Fischer’s argument appear to presuppose the falsity of incompatibilism), I do agree that

---


10 Ibid. p. 87
Fischer’s (2) seems to be mistaken. This appears to create a general problem for those who would offer deterministic FSC based criticisms of PAP. They must find some way to advance the argument without making use of the claim that the counterfactual intervener, by himself, rules out alternative possibilities. This argument needs to be modified.

In a recent paper, Fischer has conceded that his earlier argument (as regimented above) falls short and has suggested a way to modify it.\textsuperscript{11} He points out that, even though Goetz may be correct in holding that Black’s device by itself does not rule out alternative possibilities, it does not follow that the obtaining of causal determinism and the presence of Black do not \textit{together} rule out alternative possibilities.\textsuperscript{12} In fact, this could be true even if determinism rules out alternative possibilities on its own. (There may be multiple explanations for the fact that Jones is unable to do otherwise.) This insight allows Fischer to argue as follows:

1. Assume that causal determinism obtains and that the Frankfurt case of Jones and Black unfolds as above.
2. At this point in the argument, causal determinism is not assumed in itself to rule out access to alternative possibilities. (Neither is it to be supposed that Black’s presence, device, and dispositions in themselves rule out such access.)
3. Causal determinism plus Black’s presence, device, and dispositions rule out Jones’s freedom at t2 to choose otherwise.
4. If Jones is not morally responsible for choosing at t2 to vote for Obama at t3, it is not in virtue of the \textit{mere} fact that he was not free at t2 to choose otherwise.

Thus:

\begin{itemize}
\item[12] Ibid. p. 328.
\end{itemize}
5. If causal determinism rules out Jones’s moral responsibility for his choice at t2, it is *not* in virtue of its eliminating alternative possibilities (if in fact it does eliminate alternative possibilities). The key to this version of the argument appears to be the move from (3) to (4). What justifies this move? Fischer provides the following claim:

C. The fact that Black’s device (and dispositions) in a causally deterministic context rule out Jones’s freedom to choose and do otherwise is irrelevant to Jones’s moral responsibility.

C underwrites the move from 3 to 4 (and avoids Goetz’s objection to premise [2] of the first version of the argument) if we take it to be equivalent to:

C* The fact that causal determinism plus Black’s presence, device, and dispositions rule out Jones’s freedom to choose and do otherwise is irrelevant to Jones’s moral responsibility.

This version of Fischer’s argument has two virtues: it avoids begging the question against the incompatibilist and it also avoids making the dubious assumption that the presence of Black and Black’s device rule out alternative possibilities “in themselves and apart from the assumption of causal determinism.” For these reasons this argument appears to provide the best case yet made for the conclusion that deterministic FSCs show us that the mere lack of alternative possibilities is irrelevant to moral responsibility. Nevertheless, I am not convinced that it will ultimately prove to be persuasive.

Fischer’s strategy (in both versions of his argument) is to identify some feature of deterministic FSCs that eliminates alternative possibilities but is also irrelevant to the

---

13 Ibid. Note that as stated Fischer’s (2) requires us to refrain from assuming that causal determinism does by itself rule out alternative possibilities. Fischer makes it clear, however, that he thinks the argument would go through even if we were to make this assumption.

14 Fischer accepts C* as a fair restatement of C (personal correspondence with Fischer).
agent’s moral responsibility. In the first version that feature was Black. Recall premise (3) of the original formulation of Fischer’s argument:

3. Black’s presence (and device) and dispositions in themselves are irrelevant to moral responsibility.

I find (3) to be quite plausible. I think it is intuitive to hold that Black’s presence is irrelevant to Jones’s responsibility. The problem with appealing to (3), as noted above, is that it is not clear that Black’s presence, device and dispositions succeed, by themselves, in ruling out alternative possibilities. Now in the second version of the argument the feature that Fischer (in C*) identifies as being irrelevant to moral responsibility is ‘the fact that causal determinism plus Black’s presence, device, and dispositions rule out Jones’s freedom to choose and do otherwise’. I contend that when we consider what motivates us (when considering FSCs) to accept claims like (3), we will discover that this motivation does not apply to claims like C*, and thus we will be left without any strong reason to accept C* or any similar premise.

III

Fischer has helpfully pointed out that there is a distinction between the “A-Factors” of a situation, which play a role in bringing about a particular event, and the “B-Factors”, which render the event inevitable but need not cause or bring about the event.15 Drawing from this, let us say that a mere B-Factor is a B-Factor that is not also an A-Factor. Since Black plays no role in causally influencing Jones in the actual sequence of events, he is a mere B-Factor and also appears to be irrelevant to moral responsibility.

Frankfurt made a quite similar point in the paper in which he first introduced FSCs:

There may be circumstances that constitute sufficient conditions for a certain action to be performed by someone and that therefore make it impossible for the person to do otherwise, but that do not actually impel the person or in any way produce his action…An examination of situations characterized by circumstances of this sort cast doubt, I believe, on the relevance to questions of moral responsibility of the fact that a person who has done something could not have done otherwise.  

So as Frankfurt apparently saw it, there is a particular feature of FSCs which makes them cases in which the agent’s (purported) lack of the ability to do otherwise appears to be irrelevant to his moral responsibility. This feature is the fact that (in FSCs) the thing that makes it impossible for the agent to do otherwise is a mere B-Factor that is not involved in the actual production of the agents act. It seems clear to me that Frankfurt was correct about this. FSCs originally motivated us to view the ability to do otherwise as irrelevant to moral responsibility because it appeared that the thing that was eliminating alternative possibilities was a mere B-Factor.  

This suggests that we are implicitly motivated by something like the following principle:

IR: If a mere B-Factor can eliminate the ability to do otherwise then having access to alternative possibilities is not in itself relevant to being morally responsible.

I do not mean to claim that when considering FSCs we explicitly invoke IR and reason from it to the conclusion that the possessing alternative possibilities is irrelevant to moral responsibility.


17 In the article in which he introduces the distinction between A-Factors and B-Factors, Fischer appears to be amenable to this view, He says, “...in the Frankfurt cases the B-Factor is not an A-Factor; indeed, this is the signature structure of preemptive over-determination that seems to be distinctively potent in calling into question the Principle of Alternative Possibilities (PAP).” Fischer “Replies” p. 269.
responsibility. Rather I suggest that our intuitions about the relevance of an agent lacking alternative possibilities in a particular case are sensitive to whether the feature of the case that eliminates alternative possibilities is an A-Factor or a mere B-Factor, and that the way in which our intuitions are sensitive to this feature shows that we are (at least implicitly) being motivated by something like IR.

With this in mind consider again C*:

C* The fact that causal determinism plus Black’s presence, device, and dispositions rule out Jones’s freedom to choose and do otherwise is irrelevant to Jones’s moral responsibility.

It is not clear that C* is motivated by IR. This is because it is plausible that the obtaining of causal determinism is not a mere B-Factor. The question is whether causal determinism plays a role in ‘bringing about’ or ‘producing’ events in the sense relevant to Fischer and Frankfurt’s distinction. Strictly speaking what we are concerned with here is the deterministic nature of the actual sequence causal chain. The truth of causal determinism as a global fact about the world is not by itself an A-factor or a B-factor. So is the deterministic nature of the actual sequence an A-Factor which plays a role in producing Jones’s act?

In my view we should construe ‘producing’ in a broad manner that includes not just the events that cause the action but also facts about the strength of the causal

---

18 Strictly speaking it is the deterministic actual sequence causal chain that plays a role both in rendering the event inevitable and in bringing it about. The truth of causal determinism as a global fact about the world is not by itself an A-factor or a B-factor. But I will follow Fischer in referring to the deterministic actual sequence causal chain as casual determinism in this context.
relations the events stand in. Such facts are often relevant to moral responsibility.\(^{19}\) And, because of this, the distinction between A and B factors can only do its work if we construe ‘producing’ in this broader manner.

Here is an example that shows that the strength of a causal relation is sometimes relevant to moral responsibility. Suppose Bill and Bob are both pushed over by a strong wind. In Bill’s case the wind exerted a strong enough force to causally guarantee that Bill would fall no matter how much effort he made to avoid falling. But in Bob’s case the causal force of the wind was weaker. The wind’s causal influence raised the probability of Bob’s falling but did not guarantee it. It seems clear that Bill is not morally responsible for falling. But Bob may very well be responsible for falling. Here the difference with regard to responsibility in the two cases is explained by the difference in the strength of the causal relations.

So we should deem the deterministic nature of the actual sequence an A-Factor. And of course, if causal determinism is an A-Factor then “causal determinism plus Black’s presence” is also an A-Factor. And given this, C* is not supported by IR.

Now since C* is not supported by IR, it is not obvious that there is any strong motivation to accept C*.\(^{20}\) So if someone were initially inclined to think that the ability to do otherwise is relevant to moral responsibility, it is not clear why an argument that appeals to principles like C* should persuade him to reject this. Even those who are...

---

\(^{19}\) Here is another example that supports this, if the causal facts are such that my act has a 60% chance of producing a certain outcome, it is plausible that I am more responsible for the outcome than I would have been if my act had only a 30% chance of producing the outcome. In both cases my act causes the outcome. The difference in responsibility is determined by the nature of the causal relation.

\(^{20}\) Note that I do not endorse IR. I think that there are independent reasons for doubting its truth. I merely identify it as the motivating force behind our intuitions about FSCs.
agnostic about the importance of the ability to do otherwise, if they lack a strong reason to accept C*, need not be moved by deterministic FSCs into giving up their agnosticism. Thus, it appears that, in order for his argument to have significant persuasive force, Fischer needs to provide us with additional reason to accept C*.

Again the problem seems to generalize. Since Black does not rule out alternative possibilities by himself, it appears that there is no mere B-Factor which does so. Thus defenders of deterministic FSC based critiques of PAP must claim that there is a non-mere B-Factor which rules out alternative possibilities in a way that is irrelevant to moral responsibility. But since such a claim would not appear to be supported by IR, it is unclear if any strong reason to accept it could be provided.

IV

Fischer has suggested that there is a way in which C* might be indirectly supported by IR. First, note the following distinction between two types of A-factors:

**Aa-Factors**: An Aa-Factor both renders an event inevitable and deterministically brings it about, and the fact that it deterministically brings the event about is the sole reason that it renders the event inevitable.

**Ab-Factors**: An Ab-Factor both renders an event inevitable and deterministically brings it about, but it is not solely the fact that it deterministically brings the event about that renders the event inevitable.

The idea here is that Ab-Factors have components that do not contribute to bringing the event about, but that do contribute to rendering it inevitable. We can see that *determinism plus Black’s presence* is an Ab-Factor. This is because even if we grant that determinism itself renders Jones’s vote inevitable, it is not solely determinism that does this. Black’s

---

21 Fischer suggested the following line of reply to me in personal correspondence.
presence also clearly plays a partial role in rendering Jones’s vote inevitable (albeit perhaps a merely gratuitous one). With this in mind let us consider IR again:

**IR:** If a mere B-Factor can eliminate the ability to do otherwise, then having access alternative possibilities is not in itself relevant to being morally responsible.

Fischer suggests that endorsing IR should also lead us to endorse a claim like the following:

**IR2:** If an Ab-Factor can eliminate the ability to do otherwise, then having access alternative possibilities is not in itself relevant to being morally responsible.

This suggestion has some intuitive force behind it. There do seem to be important similarities between IR and IR2. Furthermore, IR2 supports C*. So endorsing IR2 would allow Fischer’s argument to go through after all. Establishing something like IR2 appears to be the right way to go for those who want to use deterministic FSCs to undermine PAP. Ultimately though, while I think that we have good reason to accept IR, I am not convinced that we also have good reason to accept IR2.

**IR** appears to be motivated by a principle about relevance. We start out with the fact that mere B-Factors are irrelevant to our possessing the kind of control required for moral responsibility. From that it apparently follows that if a B-factor can by itself eliminate the ability to do otherwise, then the lack of this ability must also be irrelevant. This chain of reasoning is supported by the following principle:

**Relevance1:** If X is irrelevant to A’s possessing the kind of control required for moral responsibility and X is sufficient by itself to guarantee the presence of Y, then Y is (in itself) irrelevant to A’s possessing the kind of control required for moral responsibility.

---

22 Black’s role may be gratuitous in the sense that causal determinism may, by itself, already rule out the ability to do otherwise.
A similar line of reasoning may seem to lead us to IR2. If we begin with the idea that a component of an Ab-Factor (i.e. a component such as Black’s presence) is irrelevant to responsibility but nevertheless contributes to the absence of alternative possibilities, it may seem to follow that the lack of alternative possibilities is also irrelevant to responsibility. This chain of reasoning, however, does not depend on Relevance1, rather it depends on the following:

Relevance2: If X is irrelevant to A’s possessing the kind of control required for moral responsibility and X contributes to the presence of Y, then Y is (in itself) irrelevant to A’s possessing the kind of control required for moral responsibility.

But I am not convinced that we should accept Relevance2. This is because it is consistent with X’s contributing to the presence of Y that Y is (in every possible case in which it occurs) also guaranteed to be present by some other factor Z that is indeed relevant to A’s responsibility. And if every time Y is present it is guaranteed to be present by a factor Z that is relevant to moral responsibility, then it does not seem obvious that the fact that X contributes to Y in some cases entails that Y is not relevant to A’s moral responsibility in these cases.

Many philosophers may have an additional reason to be skeptical of Relevance2. Since accepting Relevance2 apparently yields the result that possessing alternative possibilities is not directly relevant to moral responsibility, anyone who finds this result counterintuitive will have reason to doubt Relevance2. In my view, Relevance2 is less intuitively plausible than the claim that possessing alternative possibilities is relevant to moral responsibility. Thus, I take it that those who share my intuitions have an additional reason to doubt Relevance2.
We can perhaps make the nature of the disagreement clearer by returning to the case of Jones. There are two different ways in which Jones’s inability to do otherwise might be relevant to his moral responsibility. It might be directly relevant (i.e., it might in and of itself make a difference as to whether he is morally responsible). Alternatively, Jones’s inability to do otherwise might be relevant only as sign of some further thing (e.g., determinism) which could in turn be directly relevant to his responsibility. Fischer’s view is that (1) the fact that Black’s presence contributes to Jones’s lacking alternative possibilities entails that (2) Jones’s inability to do otherwise is not directly relevant to his responsibility (although it may still be relevant as a sign).

However, it is just not obvious to me that the mere fact that the (admittedly irrelevant) presence of Black contributes to Jones’s inability to do otherwise entails that the inability to do otherwise is irrelevant to responsibility. If Jones’s inability to do otherwise can also be completely explained by the presence of causal determinism in the actual sequence (without reference to Black at all), then it seems to me that his inability to do otherwise is plausibly directly relevant to his responsibility.

To sum up, I take it that IR is true because Relevance1 is true. But Relevance1 does not support IR2. So it is apparently open to us to accept IR while rejecting IR2. Relevance2 does support IR2, but it is not obvious that Relevance2 is true. I do not wish to claim that Relevance2 is obviously false, but I do claim that it is not unreasonable to be skeptical about Relevance2 (especially given the intuitive plausibility of PAP).

---

Ultimately it appears that Fischer’s argument (in its current form) leads us to a stalemate. Some may find IR2 to be very plausible and thus be moved to accept Fischer’s argument, but, unlike IR, it is not intuitively obvious that we should accept IR2. Thus, many philosophers will remain unconvinced that deterministic FSCs show that the lack of alternate possibilities is irrelevant to moral responsibility. It would seem that, for those who wish to advance arguments similar to Fischer’s, the strategy going forward should be to either provide arguments for Relevance2 and IR2, or alternatively, attempt to present an independent motivation for accepting C* or some similar claim. Although I cannot rule out the possibility of one of these strategies ultimately succeeding, I do take myself to have shown that utilizing deterministic FSCs to provide a convincing refutation of PAP is significantly more difficult than some may have thought.

V

I now turn to the second sort of indirect Frankfurt-style argument. This strategy attempts to draw an anti-PAP conclusion from the FSCs even while granting that the agent may possess alternative possibilities in FSCs. This strategy, like Fischer’s, can (but need not) grant the success of the dilemma defense in showing that there is no FSC in which Jones is both morally responsible and lacking alternative possibilities. But rather

---

24 One way in which one could attempt to motivate C* would be to claim that Divine foreknowledge (a) rules out alternative possibilities and (b) is a mere B-Factor. However, both of these claims are quite controversial. Furthermore, once this result was established it would superfluous to appeal to FSCs in a critique of PAP. Given IR, it follows directly from (a) and (b) that having alternative possibilities is irrelevant to moral responsibility.
than focusing on deterministic FSCs, the second strategy attempts to glean an anti-PAP conclusion from reflection on FSCs even while granting that the agent may possess alternative possibilities in the FSCs. Thus they can perhaps be construed as grappling with the indeterministic horn of the dilemma defense.

The most promising anti-PAP conclusion for this strategy to draw is virtually the same as Fischer’s. The claim is that the FSCs suggest or show that possessing alternative possibilities is (at least sometimes) irrelevant to whether or not an agent is morally responsible. Even if it turns out that any time an agent is responsible she does possess alternative possibilities, it is not in virtue of her possessing alternative possibilities that she is morally responsible. Frankfurt himself has provided an example of this indirect strategy:

. . . the usefulness of the examples . . . does not really depend upon supposing that they describe circumstances that actually make an action altogether unavoidable while playing no role in bringing the action about. The examples effectively undermine the appeal of PAP even if it is true that the circumstances that do bring an action about invariably leave open the possibility that the action might not be performed.25

But why do the examples show that alternative possibilities are irrelevant even if they do not provide a counterexample to it? By way of answer Frankfurt makes at least two distinct points. First, he suggests that the key feature of the FSCs is that they reveal that rendering an action unavoidable is not the same thing as bringing about the action.26 In short, being a B-Factor and being an A-Factor are distinct properties. Frankfurt thinks


26 Frankfurt 2003.
that recognizing this distinction should lead us to reject PAP even if no FSC ultimately provides us with a direct counterexample to it. Frankfurt reasons as follows:

Appreciating this distinction [between being a B-Factor and being an A-Factor] tends to liberate us from the natural but nonetheless erroneous supposition that it is proper to regard people as morally responsible for what they have done only if they could have done otherwise. Once it is clear that the unavoidability of an action implies nothing at all as to what actually accounted for the fact that the agent performed the action, PAP loses its otherwise rather compelling plausibility.²⁷

The problem with this line of thought, in my view, is that it does not follow from the distinctness of being a B-Factor and being an A-Factor that “the unavoidability of an action implies nothing at all as to what actually accounted for the fact that the agent performed the action.” Unless it is genuinely possible for there to be a mere B-Factor, for something to truly render an action inevitable without playing a role in bringing it about, then the unavoidability of an action does imply something quite significant about how the action was produced! Namely, that it was produced by something that rendered it unavoidable.

From the fact that being a B-Factor and being an A-Factor are distinct properties, it does not follow that something can be a B-Factor without also being an A-Factor. The FSCs originally purported to show that this was possible by providing a case in which Black was a B-Factor but not an A-Factor. However, if we do not have such a case, then we do not have a reason to accept Frankfurt’s claim about the severability of rendering an event inevitable and bringing the event about.

Frankfurt also offers the following line of reasoning:

[The agents in the FSCs’] responsibility would have been unaffected, after all, even if the fact that [the counterfactual interveners] were in place had made the actions unavoidable. Why should the presence of [the counterfactual interveners] be thought to be at all pertinent to the agents’ moral responsibility, when it played no role in leading them to do what they did and when it in no way undermined or impaired the voluntariness of their actions?28

The idea here seems to be that we can consider what would have been the case if Black had succeeded in (all on his own) ruling out Jones’s alternative possibilities. And when we consider this we should be able to see that Jones’s responsibility would have been completely unaffected. It is not so clear to me that we can see this. Since Frankfurt is granting (at least for the sake of argument) that we do not have a genuinely possible case in which Black rules out Jones’s responsibility all on his own, he is asking us to consider what would be the case if the impossible happened.

One might worry that our judgments about such cases are unreliable or that all such claims are merely trivially true and have no theoretical import. But suppose that we grant to Frankfurt that the following is a substantive truth:

IRB: If a mere B-factor did (counter possibly) eliminate the ability to do otherwise the mere B-Factor would still be irrelevant to moral responsibility.

It seems plausible that we can move from IRB to:

IRC: If a mere B-factor did (counter possibly) eliminate the ability to do otherwise then having access alternative possibilities would not be in itself relevant to being morally responsible.

However, I do not think that it is safe to move from IRC to a claim like ‘having access alternative possibilities is not in itself relevant to being morally responsible’. In general it does not follow from the fact that two things part ways in certain impossible scenarios, that those things are not necessarily or even essentially connected. Consider the following claims:

(A) If (counter possibly) there were a planet with water on it but without H2O then being partially composed of oxygen would not be an essential property of water.

(B) If (counter possibly) hatred were a virtue then the virtuous life would not be worth living.

It seems clear to me that accepting (A) and (B) should not lead us to reject claims such as ‘water is identical to H2O’ or ‘the virtuous life is not worth living. So these sorts of claims do not, in general, license strong conclusions such as the one Frankfurt wants to draw. Now it may be that there is an important difference between IRC and (A) and (B) that will allow an argument like Frankfurt’s to go through, but this would need to be argued for. Ultimately it appears that Frankfurt’s argument requires significantly more development and defense if it is to successfully undermine PAP.

One might worry that I am guilty of an over literal reading of Frankfurt. Perhaps the primary point he is making is that FSCs help us see that PAP is irrelevant. The modal
claims might then be seen as unnecessary to the larger point. What I say in the next section should provide a reply to this way of taking Frankfurt’s argument.

VI

Felipe Leon and Neal Tognazzini have recently suggested that FSCs might be used to show that possessing the ability to do otherwise is not a conceptually necessary constituent or essential property of being morally responsible. Leon and Tognazzini have in mind a notion of essential property that goes beyond an object’s merely bearing a property in all possible worlds where the object exists. Leon and Tognazzini do not commit themselves to a positive account of what it is to be an essential property. But they do suggest that definitional accounts, on which an object’s essential properties are those appealed in defining the object, are plausible. Rejecting the modal account of essential properties makes room for the view that the ability to do otherwise is not an essential property of morally responsible agency even if there are no counterexamples to PAP.

Leon and Tognazzini present an FSC which they agree does not provide a counterexample to PAP (because not all alternative possibilities are ruled out). They then comment on the case as follows:

Nevertheless, we think that this sort of case makes plausible the claim that whether there are any alternative courses of action open to [the agent] is just irrelevant to our assessment of [the agent’s] choice and action. What does seem relevant is the reason for which [the agent] acted.30


30 Leon and Tognazzini P. 562
I agree that when considering the FSCs it seems that alternative possibilities are irrelevant. But I don’t think that we should conclude they are in fact irrelevant or that the ability to do otherwise is not an essential property of moral responsible agency.

One initial point worth making is that it is not obvious that the claim that the ability to do otherwise is not an essential property of moral responsible agency entails the claim that alternative possibilities are irrelevant to responsibility. It is plausible that features that are not essential properties of X (in the definitional sense) can nonetheless explain why X is not present.\(^{31}\) Here are two plausible examples of this:

First, one might think that the fact that maximizing consequentialism is true (assuming that it is) is no part of the definition of moral permissibility. Those who reject maximizing consequentialism do not thereby fail to know what it is for an act to be permissible. But the fact that maximizing consequentialism is true is still relevant to whether particular acts are morally permissible.

Second, to borrow an example from Jonathan Schaffer, suppose that Y’s being roughly spherical is grounded in its being shape S. In this case, Y’s having shape S is clearly relevant to its being roughly spherical. But it does not seem that being shape S is a conceptually necessary constituent or essential property of being roughly spherical.\(^{32}\)

Leon and Tognazzini could grant this general point and still claim that the FSCs provide intuitive support the claim that alternative possibilities are sometimes irrelevant to moral responsibility. How should a defender of PAP respond to this? In my view the

---

\(^{31}\) This is clearly true for causal explanations. My examples are aimed at showing that it is plausible for non-causal explanations as well.

\(^{32}\) Schaffer puts this example to a different purpose in “Grounding, Transitivity, and Contrastivity” included in \textit{Grounding and Explanation}, eds. Correia and Schnieder (2012), 122-138: Cambridge
defender of PAP can grant that when considering the FSCs it seems intuitive that alternative possibilities are irrelevant to moral responsibility. This is because there is a plausible explanation of our having this intuition which is consistent with the claim that possession of alternative possibilities is always relevant to moral responsibility.

One unique feature of FSCs is that they lead us to believe that mere B-Factors can rule out alternative possibilities. The dilemma defense gives us good reason to suspect that this is an illusion. However, when considering an FSC, this is an illusion we easily fall into. Since it is plausible that mere B-Factors are irrelevant to moral responsibility, the FSCs tend to produce in us the illusion that alternative possibilities can be removed by something irrelevant to moral responsibility. This illusion plausibly explains why alternative possibilities seem irrelevant in the FSCs. Suppose that both of the following principles are essential truths about moral responsibility:

(1) Mere B-Factors are irrelevant to moral responsibility.

(2) Alternative possibilities are necessary for moral responsibility.

If this is the case then we should not be surprised that FSCs, which create the illusion that these two principles conflict, tend to produce misleading intuitions in those who consider them. To consider an FSC is to consider an impossible scenario in which one of these two putatively essential truths cannot hold. Of course our intuitions about such scenarios will be unreliable. So we have a nice explanation of why the FSCs make it seem that alternative possibilities are irrelevant to moral responsibility. We need not give up PAP to account for this seeming.
I have argued that indirect Frankfurt-style arguments are not persuasive. I do not wish to suggest that the FSCs are of no use whatsoever. Michael McKenna has suggested that the FSCs are useful because they point us towards a different conception of freedom. He calls this conception of freedom Frankfurt’s Quality of Will Thesis:

What is relevant to moral responsibility is the moral quality of the motive with which the agent acts, combined with the manner in which she brings her actions about.33

McKenna’s suggestion is that we should take the FSCs as a useful tool in bringing to our attention the sort of freedom embodied by Frankfurt’s Quality of Will Thesis. He says, “the examples, even if not decisive, help make clear a distinctive sort of freedom, one that focuses on what an agent actually does and her motive for doing it.”34 This strikes me as a useful role which the FSCs have played. The FSCs have led to the development of interesting “actual sequence” accounts of freedom.35 In my view, these accounts capture an important notion of freedom which can be used to supplement PAP in developing a comprehensive account of control. But perhaps such accounts may ultimately prove to be more plausible rivals of PAP. However, I hope to have cast doubt on the idea that their superior plausibility can be demonstrated by appeal to indirect arguments derived from the Frankfurt cases.


34 Ibid

35 In particular Fischer and Ravizza’s (1998) account.
Chapter 3: Omissions and the Frankfurt Cases: A New Challenge

The Principle of Alternative Possibilities (PAP) tells us that an agent is morally responsible for an action only if he could have done otherwise. Frankfurt-style cases (FSCs) call PAP into question because they appear to be cases in which an agent is morally responsible for an action despite the fact that he could not have done otherwise. In this essay I will argue that, despite initial appearances, we have good reason to deny that the agents in FSCs are morally responsible. As a result, we also have good reason to deny that FSCs are counterexamples to PAP. My strategy will be to begin with cases in which it is clear that an agent is not morally responsible for his behavior (or lack thereof) and argue that our judgment concerning responsibility in these cases ought to be the same as our judgment in the Frankfurt-style cases.

In 1969 Harry Frankfurt presented an extremely influential challenge to the Principle of Alternative Possibilities (PAP). This principle tells us that an agent is morally responsible for an action only if he could have done otherwise. PAP strikes most of us as initially plausible. Frankfurt, however, called PAP into question by presenting a case in which an agent appeared to be morally responsible for an action despite the fact that he could not have done otherwise. Here is a case quite similar to the one originally presented by Frankfurt:

**Original Frankfurt Case:** Black wishes Jones to cast his vote for presidential candidate A. In order to ensure that Jones does this, he implants a chip in Jones’s brain which allows him to control Jones’s behavior in the voting booth. (Jones has

---

no idea about any of this.) Black prefers that Jones vote for candidate A on his own. But if Jones starts to become inclined to vote for anyone other than A, Black will immediately use his chip to cause Jones to vote for candidate A instead. As it turns out, though, Jones votes for candidate A on his own and Black never exerts any causal influence on Jones’s behavior.

Initially it would seem that both of the following are true. (1) Jones is morally responsible for voting for candidate A. And (2) Jones could not have done otherwise than he in fact did. Thus we have an apparent counterexample to PAP.

This and similar examples have generated an enormous literature. Several different responses have been made on behalf of PAP, the most significant one being the Dilemma Defense. Proponents of the Dilemma Defense maintain that Original Frankfurt Case is crucially under described. Should we say that causal determinism obtains in the actual sequence of events leading to Jones’s vote or not? The Dilemma Defense holds that in the indeterministic version of the case the presence of the counterfactual intervener cannot remove all alternative possibilities. In appealing to the deterministic version, on the other hand, critics of PAP beg the question against those who hold that moral responsibility is incompatible with the truth of causal determinism. This is because, once we stipulate that causal determinism obtains in the actual sequence, incompatibilists will no longer find it plausible that Jones is morally responsible for voting for candidate A.  

This line of reply has hardly settled the issue. Critics of PAP have responded to the indeterministic horn by attempting to fashion more sophisticated Frankfurt Style Cases (FSCs) that do rule out any alternative possibilities robust enough to be relevant to moral responsibility without requiring the presence of determinism in the actual sequence. Attempts have also been made to appeal to the deterministic FSCs in order to show (without assuming that Jones is morally responsible) that merely lacking the ability to do otherwise is irrelevant to whether an agent is responsible. Thus, the debate continues.

Although I have significant sympathies with the defenders of the Dilemma Defense, in this essay my goal is to take the debate in a different direction by putting forth a distinct challenge for those who wish to appeal to FSCs in order to cast doubt on PAP. I will argue that (quite apart from issues concerning whether the FSCs should be construed deterministically or indeterministically) we have significant reason to doubt whether Jones is responsible. My strategy will be to begin with cases in which it is clear that an agent is not responsible for his behavior (or lack thereof) and argue that our

---


40 Here and throughout it should be assumed that by ‘responsible’ I mean morally responsible (unless I explicitly state otherwise).
judgment concerning responsibility in these cases ought to be the same as our judgment in the Frankfurt-style cases.

In section I, I will present my main argument. In section II, I will present a general challenge to anyone who wishes to reply to my main argument. In sections III-V, I provide further defense of my argument by responding to various objections that arise from the current literature on FSCs. And in section VI, I consider some methodological issues and respond to a further potential criticism.

Section I

I begin with the following case borrowed from John Martin Fischer and Mark Ravizza:

**Sharks:** John is walking along the beach and sees a child drowning in the water. John believes that he could rescue the child without much effort. Due to his laziness, he decides not to attempt to rescue the child. The child drowns. Unbeknownst to John, there is a school of sharks hidden beneath the water. If John had attempted to rescue the child, the sharks would have eaten him and his rescue attempt would have been unsuccessful.41

In this case, John is responsible for something quite important. He is responsible for not _trying_ to save the child. But, importantly for our purposes, it also seems clear that John is not morally responsible for _failing to save the child_. (I know of no one in the literature

who has claimed that he is.) So this claim will function as the first premise of my argument:

(P1) In Sharks John is not responsible for failing to save the child.

But now consider:

**Penned-in Sharks:** Everything occurs just as in Sharks except for the fact that the sharks are penned up. However, unbeknownst to John, there is an evil observer who wishes for the child to drown. If John had jumped into the water, the evil observer would have released the sharks, and as a result, the sharks would still have prevented John from rescuing the child. But the presence of the observer plays no role in the actual sequence of events.42

It seems to me that we ought to say the same thing about John’s responsibility in Penned-in Sharks as we say in Sharks. Thus:

(P2) If John is not responsible for failing to save the child in Sharks, then he is not responsible for failing to save the child in Penned-in Sharks.

Initially it is hard to see why the mere fact that the sharks are penned up but would be released if John enters the water could make a difference to John’s responsibility. Thus, I take it that anyone who would claim that (P2) is false would have some explaining to do. They would need to identify a moral responsibility-relevant difference between Sharks

---

42 This case is also from Fischer and Ravizza (1998) p. 138. They credit David Kaplan for suggesting the case.
and Penned-in Sharks. Some have attempted to do just this and these attempts will be discussed in later sections.\textsuperscript{43}

Moving on for the meantime, let us now consider a third case:

**Sloth:** In this case, there are no sharks present to prevent a rescue by John. The evil observer is now monitoring John’s thoughts instead. John decides (without deliberating much) to refrain from saving the child. If John had seriously considered attempting to rescue the child, the evil observer would have caused him to experience an irresistible urge to refrain from saving the child. However, this observer still plays no role in causing John’s decision to refrain from attempting a rescue.\textsuperscript{44}

We have now come to a case where it does seem intuitive that John is responsible for failing to save the child. Despite this, my view is that John’s responsibility in Sloth should not differ from his responsibility in Penned-in Sharks. Thus, I suggest:

(P3) If John is not responsible for failing to save the child in Penned in Sharks, then he is not responsible for failing to save the child in Sloth.

The difference between Penned-in Sharks and Sloth is that in Penned-in Sharks the counterfactual intervention would have occurred after John’s decision to try to help, while in Sloth the counterfactual intervention would have occurred before John makes a decision. It seems strange to hold that a mere difference in the timing of the

\textsuperscript{43} This line is taken by Fischer and Ravizza (1998) and Jeremy Byrd ‘Moral Responsibility and Omissions’ *The Philosophical Quarterly* 57 (2007): 56-67.

\textsuperscript{44} This sort of case was suggested by Harry Frankfurt, “An Alleged Asymmetry between Actions and Omissions.” *Ethics* 104 (April 1994): 620-623.
counterfactual intervention would result in a difference concerning John’s responsibility, but some have done so.\textsuperscript{45, 46} In particular, it has been suggested that it matters that in Sloth it is true that if John had intended to save the child, then he would have been able to do so. In section IV, I will respond to attempts to show that this difference is responsibility-relevant.

Let us now consider one last case:

**Hero:** John decides (without deliberating much) to rescue the child, and he successfully does so. Unbeknownst to him, if he had seriously considered refrain from rescuing the child, our now benevolent observer would have caused him to immediately experience an irresistible urge to rescue the child.\textsuperscript{47}

Again I see no reason to think that John’s responsibility should differ from Sloth to Hero. So I endorse:

(P4) If John is not responsible for failing to save the child in Sloth, then he is not responsible for saving the child in Hero.

I suspect that most will join me in finding (P4) to be intuitively obvious. One way to reject (P4) would be to argue, as Fischer and Ravizza once did, that the ability to do otherwise is required in order to be responsible for an omission, but not in order to be

\textsuperscript{45}Fischer and Ravizza (1998) make this critical point. See p. 140-141.


\textsuperscript{47} This case is drawn from Fischer and Ravizza, “Responsibility and Inevitability” *Ethics* 101 (1991), pp. 258-78.
responsible for an action (call this view The Asymmetry Thesis).\textsuperscript{48} Harry Frankfurt introduced the Sloth case in order to show that we should not accept The Asymmetry Thesis, and it seems clear to me that he is correct about this.\textsuperscript{49} I will defend Sloth’s status as a counterexample to The Asymmetry Thesis in section V.\textsuperscript{50}

We can now present my central argument in its entirety. (Let’s call it the No Principled Difference argument):

(P1) In Sharks John is not responsible for failing to save the child.
(P2) If John is not responsible for failing to save the child in Sharks, then he is not responsible for failing to save the child in Penned-in Sharks.
(P3) If John is not responsible for failing to save the child in Penned-in Sharks, then he is not responsible for failing to save the child in Sloth.
(P4) If John is not responsible for failing to save the child in Sloth, then he is not responsible for saving the child in Hero.

Thus;

(Conclusion) John is not responsible for saving the child in Hero.

Now since Hero is structurally identical to Original Frankfurt Case, the cogency of this argument gives us reason to doubt that Jones is responsible for voting as he does. And of course Original Frankfurt Case is only a counterexample to PAP if Jones is in fact responsible in Original Frankfurt Case. So the challenge to those who wish to wield FSCs

\textsuperscript{48} Fischer and Ravizza (1991).

\textsuperscript{49} Frankfurt 1994.

\textsuperscript{50} Carolina Sartorio argues that Sloth is not a counterexample to The Asymmetry Thesis in ‘A New Asymmetry Between Actions and Omissions.’ Noûs 39 (2005), pp. 460-482.
in a critique of PAP is to show that we can draw a principled line somewhere between Sharks and Hero such that we can plausibly maintain that John is responsible in Hero but not in Sharks. Accomplishing this would of course require rejecting (P2), (P3) or (P4). I will now argue that that there is good reason to think that it is in principle quite difficult for this challenge to be met and that no one has as of yet successfully met it.

Section 2

The paper in which Frankfurt first introduced FSCs contains the following remarks on the unique feature of FSCs in virtue of which they appear to be counterexamples to PAP:

There may be circumstances that constitute sufficient conditions for a certain action to be performed by someone and that therefore make it impossible for the person to do otherwise, but that do not actually impel the person or in any way produce his action…An examination of situations characterized by circumstances of this sort cast doubt, I believe, on the relevance to questions of moral responsibility of the fact that a person who has done something could not have done otherwise.\footnote{Frankfurt (1969).}

In a similar vein, Fischer has pointed us to an important distinction between the “A-Factors” of a situation, which bring about a particular event, and the “B-Factors” which render the event inevitable but need not cause or bring about the event.\footnote{Fischer, “Replies” \textit{Philosophy and Phenomenological Research} (Vol. LXXX No. 1, January 2010):267-278. See p. 269.}
Fischer’s terminology with Frankfurt’s suggestion yields the following account. The reason it is so intuitively plausible that responsibility is not undermined in the FSCs is that FSCs are purportedly cases in which the factors that remove the agent’s ability to do otherwise are mere B-Factors, and mere B-Factors appear to be irrelevant to whether an agent has the type of control he needs in order to be moral responsible. The presence of Black in Original Frankfurt Case, for example, appears to make it impossible for Jones to avoid voting for candidate A without actually causing Jones to vote for A.

I think that Frankfurt has correctly identified the feature of the FSCs in virtue of which they appear to be counterexamples to PAP. To see this, imagine variations on the FSCs in which the intervener does play an active causal role in bringing about the agent’s act (and is thus not a mere B-Factor). In these variants it becomes much less intuitive to hold that the agent’s responsibility is unaffected by the presence of the intervener. The reason Black appears to be irrelevant in Original Frankfurt Case is precisely because he does not make anything happen in the actual sequence of events. Thus it appears that the principle underlying our intuitions about FSCs is that mere B-Factors are irrelevant to moral responsibility.

This leads to an additional challenge for anyone who wishes to reply to the No Principled Difference argument. The problem is that each case appealed to in the No Principled Difference argument centrally involves the presence of a mere B-Factor (the sharks in Sharks, the evil observer in Penned-in Sharks, etc.). Furthermore, accepting that

53 ‘mere’ in the sense that they are not also A-Factors.

54 Given this picture, one way to understand the dilemma defense is that it calls into question whether it is true that a mere B-Factor can rule out the ability to do otherwise.
the agent is not responsible in any of the cases apparently involves rejecting the claim that mere B-Factors are always irrelevant to moral responsibility. Since defenders of FSCs should say that the principle underlying our intuitions about FSCs is correct, it will apparently be quite difficult for them to accept the claim that John is not responsible in any of the cases appealed to in the No Principled Difference argument. Thus, it will be difficult for them to draw a line anywhere between Sharks and Hero with regard to John’s responsibility.

There are several potential lines of reply to this point. One would be to claim that the presence of the sharks (in Sharks) does play a causal role in the actual sequence and thus is not a mere B-Factor. The difficulty with this reply is that it is quite counterintuitive to hold that the sharks cause you to fail to save the child. (In section IV I will discuss a variant of this reply on which the presence of the sharks is said to indirectly affect the causal structure of the sequence by preventing John’s omission from causing the child’s death.)

A second line of reply would be to reject the claim that it is because Black is a mere B-Factor that it seems to us that he is irrelevant to moral responsibility. How plausible this reply would be depends to some degree on the details of the rival account, but it does seem that there is a significant cost to rejecting this claim. Again, it seems that Black is irrelevant precisely because he does not make anything happen in the actual sequence of events.

There is a version of this second line which I do find somewhat plausible. One could deny (P3) on the grounds that the Dilemma Defense is successful and that in
(indeterministic versions of) Sloth and Hero the counterfactual intervener fails to remove
John’s ability to save the child. On this view, Black is irrelevant in Original Frankfurt
case, not because he fails to cause anything, but rather because his presence does not
remove alternative possibilities (or affect anything else Jones needs in order to be
responsible). Other mere B-Factors, such as the sharks (in Sharks) do remove some of
John’s options, and these are relevant to responsibility. The problem with this view is that
it involves accepting that Jones is responsible precisely because he has alternative
possibilities: thus rendering FSCs useless with regard to their original purpose. How
plausible this account turns out to be will depend to a great extent on whether the
Dilemma Defense is ultimately successful. The important thing to note for our purposes
is that this reply is not open to anyone who wishes to use FSCs to attack PAP.

In what follows I will discuss particular accounts in the literature which, if
correct, would license a rejection one of the premises of the No Principled Difference
Argument. The upshot of the current section is that even if the reader finds the objections
I raise to one of these particular accounts unpersuasive, each of these views still faces the
general challenge of accounting for the fact that it seems that Black is irrelevant because
he is a mere B-Factor. The nature of the subject matter in the following sections will
require that the discussion become quite intricate at certain points. One of my goals in
the first two sections was to lay out the main thrust of my challenge to FSC-inspired
critiques of PAP in a relatively simple fashion. Thus, I hope that I have already
succeeded in showing (independently of what follows) that there is indeed a challenge for
FSC based critiques of PAP and that there is reason to think that it will prove difficult to meet.

Section 3

In addition to providing many of the examples I have appealed to, Fischer and Ravizza have presented a detailed account of responsibility which purports to provide a motivation for rejecting (P2). Appealing to FSCS, they suggest that, just as we must (on their view) “hold fixed” the nonintervention of Black in evaluating Jones’s responsibility for his action, so too must we hold fixed the nonoccurrence of some events in evaluating an agent’s responsibility for an outcome. To answer the question of which events must be held fixed they introduce the notion of a triggering event. A triggering event “(relevant to some consequence C) [is an] event which is such that, if it were to occur, it would initiate a causal sequence leading to C.”

For example, Black’s use of his device to cause Jones to vote for A would count as a triggering event relative to the consequence that a vote is cast for A. Fischer and Ravizza then provide us with the following necessary condition for an agent’s being responsible for the consequences of an omission, such as John’s failing to save the child:

Suppose that in the actual world an agent S moves his body in way B at time T via a type of mechanism M, and S’s moving his body in way B at time T causes some consequence-universal C to obtain at T+i via a type of process P…[Then S is only responsible for C on the condition that] If S were to move his body in way B* [which cannot be identical to B] at T, and all other triggering events (apart from

---

55 Fischer and Ravizza (1998) p. 110-111
that do not actually occur between \( T \) and \( T+i \) were not to occur, and a \( P \)-type process were to occur, then \( C \) would not occur.\(^{56}\)

Fischer and Ravizza argue that we can use this requirement to show that John can be responsible for failing to save the child in Penned-in Sharks even though he cannot be responsible for this failure in Sharks. In Penned-in Sharks the evil observer’s release of the sharks from the pen counts as a triggering event and so (given Fischer and Ravizza’s account) we should hold fixed its nonoccurrence. Thus we get the result that (assuming that John meets all other requirements for being morally responsible) John is responsible for the fact that the child drowned. And this would apparently entail (in this context) that John is responsible for failing to save the child.

Fischer and Ravizza want to maintain that their account yields the result that John is not responsible in Sharks. But it is unclear how they can get this result since it would appear that the shark’s sensing that John entered the water should count as a triggering event as well. Fischer and Ravizza discuss this worry in a footnote and they say the following:

…in the alternate sequence, John’s jumping into the water would antedate and lead to the shark’s sensing that he had done so: thus, the shark’s sensing John would not “initiate” – in the relevant sense – the sequence leading to the child’s not being saved by John (and thus would not be a triggering event).\(^{57}\)

\(^{56}\) Fischer and Ravizza (1998) p. 112 (also p. 135). Note that I have omitted to state parts of Fischer and Ravizza’s account which do not concern us here but which are important for evaluating their account in other contexts.

\(^{57}\) Fischer and Ravizza (1998). p. 136
I do not see how this reply will help (at least with the goal of distinguishing between Sharks and Penned-in Sharks in mind). This is because if the sharks’ sensing that John jumped into the water does not count as a triggering event because it is antedated and caused by John’s act, then surely the evil observer’s (in Penned-in Sharks) sensing that John jumped into the water would not count as a triggering event for the same reason. Thus, it does not appear that Fischer and Ravizza’s view (in its current form) can account for the purported difference between Sharks and Penned-in Sharks.\(^{58}\)

Jeremy Byrd has recognized that Fischer and Ravizza’s unmodified account fails to license a rejection of claims like (P2) and he suggests a modification that would do the trick.\(^{59}\) Byrd’s view is that we should only hold fixed the nonoccurrence of triggering events that are the choices of rational agents. Thus, we hold fixed the nonoccurrence of our evil observer’s decision to let the sharks out of the pen (in Penned-in Sharks). But we do not hold fixed the nonoccurrence of the sharks sensing and attacking John (in Sharks).\(^{60}\) (Byrd apparently has in mind a construal of rationality on which sharks do not count as rational agents.)

I do not think Byrd’s suggestion is satisfactory. Byrd claims that this view “provides a systematic solution which matches one’s judgments in the clear cases and

---

\(^{58}\) For a similar critique of Fischer and Ravizza’s view, see Byrd (2004).

\(^{59}\) Strictly speaking, Byrd’s view seeks to rescue Fischer and Ravizza’s account from a parallel problem having to do with responsibility for consequences. And his modified notion of triggering events is thus intended to apply only to responsibility for consequences. However, in order to see whether it can provide Fischer and Ravizza with a way to reject (P2), we can consider a modified version that applies to omissions as well.

\(^{60}\) Byrd (2004).
gives proper guidance in the tougher ones."\textsuperscript{61} However, I think that there are clear cases in which Byrd’s view yields the intuitively wrong results. Consider for example:

**Non-Agential Sloth:** In this case, there are no sharks or evil observers present to prevent a rescue by John. John is afflicted by a phobia of water of which he is completely unaware and not responsible for possessing. John decides (without deliberating much) to refrain from saving the child. If John had seriously considered attempting to rescue the child, his phobia would have caused him to experience an irresistible urge to refrain from saving the child. However, this phobia plays no role in causing John’s decision to refrain from attempting a rescue.\textsuperscript{62}

This case seems to me to call for the same verdict as the original Sloth case. I would be very surprised if many philosophers had the intuition that John is responsible in Sloth but not in Non-Agential Sloth, yet this is just what Byrd’s account appears to suggest. Since it would appear that we should hold fixed the nonoccurrence of the evil observer’s choices but not the nonoccurrence of the irresistible urge.\textsuperscript{63}

I do not wish to overstate my conclusion here; I fully admit that it is possible that a modified version of Fischer and Ravizza’s account would be able to distinguish between Sharks and Penned in Sharks. My claim is merely that such a modification

\textsuperscript{61} Byrd (2004). p. 63

\textsuperscript{62} This non-agential version is closer to how Frankfurt (1994) originally presented the Sloth case.

\textsuperscript{63} Jeremy Byrd has pointed out that his requirement that the triggering event be the choice of a rational agent could be limited to cases in which the triggering event occurs post choice. This version of the requirement would avoid yielding an incorrect verdict in Non-Agential Sloth, but at the cost of apparent arbitrariness. Furthermore, the revised requirement would still be vulnerable to a Non-Agential version of Penned-In Sharks.
would have to be presented before Fischer and Ravizza’s account could ground a reply to
the No Principled Difference argument and I, at least, see no obvious way of making such
a modification. Furthermore, even if this sort of modification were developed, Fischer
and Ravizza’s account could still be faulted on the grounds that it does not do justice to
the fact that Black is irrelevant because he is a mere B-Factor.

Section IV

Having criticized Fischer and Ravizza’s attempt to reject (P2), I will now consider
an account that, if true, would license a rejection of (P3).

(P3) If John is not responsible for failing to save the child in Penned in Sharks,
then he is not responsible for failing to save the child in Sloth.

Randolph Clarke has defended a view on which (P3) will turn out to be false. He accepts
the following necessary condition for an agent’s being responsible for an omission:

INTAB An agent is responsible for omitting to A only if, had the agent intended
to A, he would have been able to A.64

Clarke maintains that John is not responsible in Sharks and Penned-in Sharks but is
responsible in Sloth. INTAB rules out John’s being responsible in Sharks and Penned-in
Sharks, but allows that John might be responsible in Sloth. This is because in Sloth the
counterfactual intervention is set to occur before John forms the intention to save the
child. Clarke endorses INTAB primarily on the basis that it gets the right results in these
and similar cases. I agree with Clarke that INTAB yields intuitively correct results in a

64 Clarke (forthcoming) p. 25. McIntyre (1994) defends a similar view. Note that Clarke thinks that INTAB
is true only in an “appropriately restricted, revised and refined form.” None of these restrictions, etc. will
be relevant to my discussion of INTAB.
range of omissions cases. However, I think that it is ultimately implausible for several reasons.

First, as Clarke acknowledges, the truth of \textbf{INTAB} would appear to leave us with a significant asymmetry between actions and omissions. The problem is that there does not appear to be any requirement similar to \textbf{INTAB} that holds true in the case of action. The parallel to \textbf{INTAB} in the case of action would be:

\textbf{ACTION INTAB} An agent is responsible for A-ing only if, had the agent intended to refrain from A-ing, he would have been able to refrain from A-ing.

But this principle appears to be undermined by FSCs in which the counterfactual intervener is prepared to intervene immediately after the intention to refrain is formed.\footnote{Fischer points this out in 'Responsibility and the Kinds of Freedom.' \textit{Journal of Ethics} (2008). 12:203-228}

Consider:

\textbf{Post Intention Frankfurt Case}: Black wishes Jones to cast his vote for presidential candidate A. In order to ensure that Jones does this, he implants a chip in Jones’s brain which allows him to control Jones’s behavior in the voting booth. (Jones has no idea about any of this.) Black prefers that Jones vote for candidate A on his own. But if Jones forms the intention to vote for anyone other than A, Black will immediately use his chip to cause Jones to vote for candidate A instead. As it turns out, though, Jones votes for candidate A on his own and Black never exerts any causal influence on Jones’s behavior.

This case appears to show that \textbf{ACTION INTAB} is false.\footnote{At least by the lights of those who accept our basic intuitions concerning FSCs.}
Clarke is sensitive to this issue. He notes that “INTAB imposes a requirement concerning ability that, if Frankfurt is right, has no parallel in the case of action.” So accepting INTAB results (by Frankfurtian lights) in a significant asymmetry between what is required in order to be responsible for an action and what is required in order to be responsible for an omission.

I do not find this view very plausible. To see why, consider what Clarke’s account would say about Post Intention Frankfurt Case. Since we are rejecting ACTION INTAB we will say that Jones is responsible for voting for candidate A. Now suppose that in casting his vote for A, Jones thereby refrained from voting for candidate B. It seems clear to me that in this case Jones’ responsibility for voting for A and his responsibility for refraining from voting for B should stand or fall together. If Jones is responsible for voting for A, he is thereby responsible for refraining from voting for B. But if we accept INTAB we cannot say that Jones is responsible for refraining from voting for B, since it is not true that if Jones had intended to vote for B, he would have been able to.

Clarke has pointed out to me that Jones’ responsibility for voting for A and his responsibility for refraining from voting for B will not stand or fall together if Jones is unaware that he has the option of voting for B. So, Clarke suggests, it is open to him to say that responsibility for an action and a corresponding omission will not stand or fall together for other reasons as well. I agree that in some cases they will not stand or fall together. What I wish to maintain is that is implausible to hold that they do not stand or

---

67 Clarke (forthcoming) p. 39

68 Conversation with Randolph Clarke. Thanks to Clarke for helpful discussion here.
fall together in cases where the agent meets the following two conditions: (1) the agent
knows that by voting for A he is thereby refraining from voting for B and (2) the agent
either has both the ability to vote for B and the ability to refrain from voting for A or
lacks both the ability to vote for B and the ability to refrain from voting for A. Both of
these conditions can be met in a suitably filled out version of Post Intention Frankfurt
Case.

Clarke attempts to lessen the implausibility of positing this asymmetry between
responsibility for actions and responsibility for omissions by suggesting that omissions
may not be events (as actions are), but instead are the absences of certain actions. He then
reasons as follows:

If such a view is correct, it is to be expected that there might well be major
differences between what is required for responsibility for actions and what is
required for responsibility for omissions, for an omission, unlike an action, is not
an agent’s exercising control in bringing something about. Responsibility for
omissions will not be a special case of responsibility for actions, for omission
isn’t a special case of action (anymore than the absence of a lion is a special kind
of animal or the absence of red is a special kind of color). 69

While this may help some, since it creates more separation between action and omission.
It does not, at least in my view, remove the unsatisfactory arbitrariness of the asymmetry.
It still appears that in some cases being responsible for an action can entail being
responsible for a certain corresponding omission. And it seems to me that Post Intention

69 Clarke (forthcoming) p. 40
Frankfurt case is one of these cases. Furthermore, even granting Clarke’s view of the metaphysics of omissions, we still lack an explanation for why the metaphysical differences between actions and omissions result in different control requirements for moral responsibility.

In defense of his account Clarke notes that it may be appropriate to sacrifice some degree of symmetry in our theory of moral responsibility in order to account for our intuitions about different cases. I think that this correct. However, I think that we ought to regard a lack of symmetry as a cost, especially when we cannot see why the asymmetry holds. Furthermore, I do not think Clarke’s view accounts for all of our intuitions about cases, since it fails to yield the result that (in Post Intention Frankfurt Case) if Jones is responsible for voting for A, then he is responsible for refraining from voting for B.\(^{70}\)

A distinct worry for Clarke’s approach is that it yields the result that some mere B-Factors are relevant to moral responsibility. The presence of the evil observer in Penned-in Sharks is apparently a mere B-Factor, but, given INTAB, it is not irrelevant to responsibility. One could claim that INTAB provides us with an explanation for why some B-Factors are relevant while others are not. But this appears inadequate. The FSCs drive us to think that the effect various factors have on which counterfactuals are true of us is irrelevant to our moral responsibility, when those factors do not contribute at all to the actual sequence of events. If we accept this intuition, then it seems that we should

---

\(^{70}\) Clarke has pointed out to me that that this claim is plausibly construed as resulting from a more general theoretical commitment rather than as a basic intuition. This may be true. But in that case the significant plausibility of the claim suggests that the underlying theoretical commitment is one that we should not be quick to reject.
conclude that the presence of the evil observer is likewise irrelevant to moral responsibility.

Carolina Sartorio has defended a set of claims about responsibility which, if true, may allow Clarke’s view to surmount some of the objections I have raised to his account. First, Sartorio endorses the following:

\[ \text{TR(Causal): An Agent’s responsibility for } X \text{ transmits to an outcome } Y \text{ iff } X \text{ causes } Y \text{ (and the usual provisos obtain).}^{71} \]

She also argues for:

\[ \text{NA (Causal): An action can cause an outcome even if the outcome would still have occurred in the absence of the action. By contrast an omission cannot cause an outcome if the outcome would have still occurred in the absence of the omission.}^{72} \]

TR(Causal) and NA(Causal) together entail:

\[ \text{NA: An agent’s responsibility for an action can transmit to an outcome even if the outcome would have occurred anyway in the absence of the action. However, an agent’s responsibility for an omission cannot transmit to an outcome if the outcome would have occurred anyway in the absence of the omission.}^{73} \]

If NA is correct, then Clarke may be right in holding that John is not responsible in Penned-in Sharks. Since it is plausible that John does not omit to save the child if he tries

---

72 Ibid. p. 470
73 Ibid.
and fails, it is true that in the absence of the omission the child would still have drowned. And since I take it that John is responsible for failing to save the child only if he is responsible for the child’s drowning, John would not be responsible for failing to save the child.

Furthermore, we could now explain why INTAB could be true even though ACTION INTAB is false. The asymmetry is explained by the fact that actions can cause outcomes under conditions in which omissions cannot cause outcomes. And perhaps most significantly, we could also now offer a more satisfying explanation for why some B-Factors are relevant to moral responsibility. A-Factors can be relevant to responsibility because they cause something to occur in the actual sequence. B-Factors can be relevant because they prevent omissions from causing events in the actual sequence. The presence of the intervener in Penned-in Sharks does make a difference to what happens in the actual sequence of events because he prevents John’s omission from being a cause of the child’s drowning. In Sloth, by contrast, John’s omission can be a cause of the child’s drowning because it is not true that the child would still have drowned in the absence of the omission.\[^74]\ Thus the rejection of (P3) would appear to be vindicated.

Unfortunately, however, there is a compelling case to be made against endorsing NA and using it in a reply to the No Principled Difference Argument. Consider a case similar to one suggested by Frankfurt:

\[^74]\ This is because in Sloth the evil observer stops John from rescuing the child by getting him to omit to rescue the child. It is somewhat surprising that Sartorio’s view has this consequence since she is skeptical of John’s responsibility in Sloth (see section V).
No Movement: Stanley sits in a room and deliberately refrains from moving his body at all for a period of five minutes. Since Stanley is the only person in the room, his lack of movement entails that no movement occurs in the room. Unbeknownst to Stanley, there is a counterfactual intervener who would have prevented Stanley from moving his body at all had he attempted to do so.\(^75\)

It seems true in this case both that (1) Stanley caused it to be the case that no movement occurred in the room and that (2) Stanley is responsible for the result that no movement occurred in the room. However, it is not true that in the absence of the omission, the result would not have occurred. This is because if Stanley had tried and failed to move he would not have omitted to move but it would still have been true that no movement occurred in the room.\(^76\) Thus, it appears that, given that our intuitions about FSCs are correct, we have a counterexample to NA. (Note that I do not wish to claim that Stanley is in fact responsible. What I am claiming is that those who accept our intuitions about FSCs should say that Stanley is responsible.)

Here is another case that appears to be a counterexample to NA:

Lightning Sloth: There are no sharks present to prevent a rescue by John. The evil observer is now monitoring John’s thoughts. John decides (without deliberating much) to refrain from saving the child. If John had seriously considered attempting to rescue the child, the evil observer would have caused

\(^{75}\)See Frankfurt (1994) p. 620-621

\(^{76}\)One could reject this claim and hold that one omits to do A if one tries and fails to do A. However, on this understanding of omissions NA could no longer be used to rule out John’s responsibility in Penned in Sharks. This is because it would then be the case that in Penned in Sharks John still omits to save the child if he is prevented from doing so by the intervener.
him to be immediately struck dead by a bolt of lightning. However, the observer
plays no role in causing John’s decision to refrain from attempting a rescue.
In this case it is implausible to hold that the child would not have drowned had John’s
omission not obtained. It would be very strange to say that John would have omitted to
save the child if he had been struck by lightning prior to making his decision about
whether to jump into the water. NA would thus entail that John is not responsible for the
resulting death of the child. But this seems counterintuitive.

In my view it seems that Stanley is responsible in No Movement and that John is
responsible in Lightning Sloth for the same reasons that it seems that Jones is responsible
in Original Frankfurt Case. Thus, NA apparently does not successfully capture the
intuitive difference between cases where the counterfactual intervener appears to be
relevant and the cases in which he appears to be irrelevant. As a result, those who accept
our basic intuitions about FSCs ought to reject NA and thus cannot appeal to NA in order
to reject (P3).

Section V

Recall (P4):

If John is not responsible for failing to save the child in Sloth, then he is not
responsible for saving the child in Hero.

In my view (P4) is more intuitively obvious than (P2) and (P3). So I am not inclined to
think that a reply to the No Principled Difference argument which relies on rejecting (P4)
will be the best way to go. Nevertheless Sartorio has argued that we have reasons to
doctor that John is responsible in Sloth which do not also provide reasons to doubt that
John is responsible in Hero. On Sartorio’s view, the reason we think that John is responsible in Sloth is that we think that (a) he is responsible for his decision not to jump into the water and (b) his decision caused the child’s death.\textsuperscript{77}

Sartorio finds (b) to be questionable. She suggests that, “what caused the child’s death is not my decision not to jump in, but my failure to decide to jump in”.\textsuperscript{78} Now it might seem that this does not matter very much, since we could then say that (a’) John is responsible for his failure to decide to jump in and (b’) his failure to jump in caused the death of the child.\textsuperscript{79} But Sartorio also casts doubt on (a’). She points out that you cannot infer from the mere fact that no one forced John to fail to decide to jump in that he is responsible for failing to decide to jump in.\textsuperscript{80} Since both (b) and (a’) are questionable, Sartorio concludes that we are left without a good reason to think that John is responsible.

In my view we do have good reason to accept both (b) and (a’). Consider first (a’). I agree that the mere fact that John was not forced to fail to decide to jump in does not entail (a’). But in my view, (a’) is quite intuitive. Note that it is very strange to hold that John is responsible for his decision not to jump in but is not responsible for failing to decide to jump in. In normal cases responsibility for the decision and the corresponding failing appear to stand or fall together (at least when we pick out the act that the agent decided not to do and the corresponding failing under the same description). In addition I do think that there is a fairly strong argument for (a’). Consider:

\textsuperscript{77} Sartorio (2005) p. 463-464. My (a) and (b) correspond to Sartorio’s (1) and (2’)
\textsuperscript{78} Ibid. p. 464
\textsuperscript{79} My (a’) and (b’) correspond to Sartorio’s (1’) and (2’’)
\textsuperscript{80} Sartorio (2005) p. 465
(c) John is responsible for causing his failure to decide to jump into the water.

And;

(d) If John is responsible for causing his failure to decide to jump into the water then he is responsible for his failure to decide to jump into the water.\(^8\)

Now (c) and (d) together entail (a’). Both (c) and (d) appear to be quite plausible. Although in general one need not be the cause of one’s failure to decide, It seems plausible to say in this case that by deciding not to jump John causes himself to fail to decide to jump in. Furthermore, so long as we are granting that the mere presence of the evil observer does not remove responsibility for decisions, it is hard to see why John would not be responsible for causing himself to fail to decide to jump in.

One could worry that in order to say that John causes his failure to decide, I must posit simultaneous causation since John might both decide not to jump in and fail to decide jump in at the same moment. I see two ways or resolving this short of embracing simultaneous causation. One would be to claim that while John’s decision at t does not cause his failure to decide to jump in at t, it does cause his failure to decide to jump in at t+1. And John’s failure to decide at t+1 will be part of the causal explanation of the child’s death. The other option would be to couch the argument in terms of explanation rather than causation. Explanation is a broader notion than causation, and there does not appear to be any problem with simultaneous explanation. (For example, certain chemical properties of my coffee cup at t explain why my cup is hot at t.) We would then claim

---

\(^8\) Note that this is just a claim about the Sloth case. It does not purport to be a principle about responsibility in general.
that John’s deciding not to jump in provides an explanation of his failure to decide to jump in and that this grounds his responsibility for his failure to decide to jump in.

In addition to (a’), (b) is also plausible. If I am right that by deciding not to jump in the water John causes himself to fail to decide to jump in, then, given that this failure causes the child’s death, it would certainly appear that John’s decision does play a causal role in bringing about the death of the child. (Again, we could move to the claim that John’s decision plays an explanatory role if that is preferable.) So in short, (a’) and (b) both appear to be true, and if either of them are true, then Sartorio’s argument fails.

As with the other replies we have considered, those who would reject (P4) would need to explain why some mere B-Factors are relevant to moral responsibility while others are not. Why is the counterfactual intervener relevant to John’s responsibility in Sloth, but not in Hero? Presumably Sartorio would attempt to appeal to NA or some similar principle in order to do this. But as we have seen NA is not very plausible.

Section VI

If the reader has been persuaded by what I have said up to this point, then he will be convinced that none of the replies to the No Principled Difference argument currently present in the literature are successful. Each reply we considered had its own problems. In addition, we saw that there is a general problem. Most potential replies to the No Principled Difference argument would have to accept the claim that mere B-Factors are sometimes relevant to moral responsibility, thus rejecting the principle that appears to underlie our intuitive response to FSCs. However, there is a potential reply that avoids
this result. One could bite the bullet and claim that John is morally responsible for failing to save the child in Sharks. It is to this possible reply that I now turn.

The trouble generated by the No Principled Difference argument is structurally similar to issues that arise in other philosophical contexts. Consider, for example, the trolley problem. We find that there are two cases in which we are inclined to different judgments about the permissibility of an act. We judge that it is permissible to throw a switch in order to redirect a runaway trolley so that it will run over one rather than five. But we also judge that it is not permissible to kill one healthy person in order to harvest his organs and save five lives.\(^{82}\) Further reflection on related cases shows us that it is very difficult to find a principled difference which would justify our holding different views about the two cases.\(^{83}\) Someone convinced that there is no difference with regard to permissibility between the two cases might offer an argument which begins with the trolley case and concludes that it is permissible after all to kill the healthy person in the organ transplant case. However, it is not obvious that this argument would be any better than one which began with the organ transplant case and concluded that it is not permissible to throw the switch in the original trolley case.

This leads us to a potential strategy which a critic of the No Principled Difference argument might employ. This critic could reverse the direction of the argument. Once we see that there is no responsibility relevant difference between Sharks and Original


Frankfurt Case, we ought to conclude, not that Jones is not responsible in Original Frankfurt Case, but rather that John is responsible in Sharks. This new argument would still make use of the same (P2), (P3) and (P4) as my original argument. In short, our critic claims, all I have shown is that we must reject either our intuition about Sharks or our intuition about Original Frankfurt Case. I have not shown it is preferable to reject our intuition about Sharks, and thus, have not shown that critiques of PAP based on appeals to FSCs are unsuccessful.

I agree that this reply is an option for those who wish to avoid the conclusion of the No Principled Difference argument, so let me explain why I think we have more reason to hold onto our intuition about Sharks than we do to hold onto our intuition about Original Frankfurt Case. First, my intuition that John is not responsible in Sharks is initially somewhat stronger than my intuition that Jones is responsible in Original Frankfurt Case. I take it that those who agree with me about the initial weight of the two intuitions will have reason to prefer holding onto their intuition about Sharks. Second, Sharks is a more realistic, down to earth scenario than Original Frankfurt Case. In my view, all other things being equal, we have more reason to trust our intuitions about realistic cases than we do about farfetched ones. I realize that not everyone will agree with this methodological claim and I do not deny that farfetched cases are evidentially relevant. So I grant that this point may be of limited significance.

Third, my confidence that Jones is responsible in Original Frankfurt Case is diminished when I consider the following question raised by David Widerker: “what, in
your opinion, should [Jones] have done instead?" When I consider Widerker’s question I do not lose the intuition that Jones is responsible, but I do find myself feeling less confident that he is responsible. I know of no similar way of reducing my confidence that John is not responsible in Sharks. Thus I suspect that those who share my reaction to Widerker’s question will find it more palatable to reject their intuition about Original Frankfurt Case.

Lastly, if defenders of FSC based critiques of PAP are correct in holding that we should not accept PAP if our intuitions about FSCs are correct, then the intuitiveness of PAP itself provides us with good reason to prefer our intuition about Sharks over our intuition about Original Frankfurt Case. The following principle underlies this claim: if we must reject either A or B and A is consistent with intuitively plausible principle C but B is not, then, all things being equal, we ought to reject B. Many find PAP to be very intuitive; it appears that those who do will have reason to hold on to their intuition about Sharks.

In conclusion, I have argued that no one has yet successfully shown that there is a principled distinction to be drawn between Frankfurt-style cases and cases in which the agent is clearly not responsible. Furthermore, there are general reasons for thinking that this task will not be easily accomplished. This result gives us reason to doubt that our

---


85 John Perry has suggested to me that one might argue that John is partially responsible in Sharks by endorsing the following principle: “In general, if x would have done A, had he tried, but didn’t even try, x is wholly responsible for not A-ing. If he wouldn’t have succeeded, even if he tried, then x is partly responsible for failing to A.” (Correspondence with Perry.) I doubt that this principle is correct. In my view, the more natural thing to say concerning cases like Sharks is that John is wholly responsible for not *trying* to save the child, but is not responsible at all for failing to save the child.
intuitions about FSCs are reliable, and thus gives us reason to be skeptical of FSC based critiques of the principle of alternative possibilities.
Chapter 4: Control, Blame and the Reactive Attitudes

“…the justification of anger, resentment and indignation seems to me to include a condition of freedom… I must admit, however, that I have not offered any account of why freedom is a condition of anger and resentment. Indeed, I am not even sure what such an account would look like.”
-Susan Wolf

Volitionalists hold that blame is only appropriate if the agent being blamed acted voluntarily. Volitionalists face the difficult challenge of explaining why the negative reactive attitudes must be morally inappropriate if the agent acted involuntarily. This problem is heightened when we compare the reactive attitudes to other negative mental states, such as distrust, which can be appropriate even if the agent lacked control. In this essay I respond to this problem by developing two plausible views of the content of the reactive attitudes which explain why control is required. In section I, I lay out the challenge to volitionalism. In section II, I propose two volitionalist accounts of guilt. In section II I expand these accounts to the other negative reactive attitudes and explain why the accounts support volitionalism. In Section IV, I respond to some objections drawn from the work of Pamela Hieronymi.

Let me begin by explaining what I mean by ‘blame’. First, I distinguish between the judgment that someone is blameworthy and the emotional and behavioral blaming responses. In my view both the judgments and the responses are sometimes picked out by ‘blame’ in ordinary language. But we should not make the mistake of failing to distinguish them. Second, there are two very different sorts of emotional and behavioral responses that various philosophers have wanted to call blame. On the one hand there is
what Susan Wolf has called *angry blame*. To blame someone in this sense is to have certain negative emotional attitudes towards him (such as resentment, indignation or guilt) and to engage in certain behaviors naturally motivated by those attitudes. On the other hand we have what Wolf has dubbed *Scanlonian blame*. Scanlonian blame is “an adjustment in attitudes toward the object of blame that appropriately reflect the impairment [of relationship] that the person blamed has indicated by his behavior.”

My primary interest is in angry blame. I think it is clear that Scanlonian blame can be appropriate in cases where the agent’s behavior was involuntary. For example, I can appropriately adjust my attitudes towards someone who compulsively lies to me in such a way that I no longer trust them. The main issue I want to explore is whether angry blame (and in particular the reactive attitudes of indignation, resentment and guilt) are appropriate in cases where the agent did not act voluntarily. So for the rest of this essay I will use ‘blame’ to refer to angry blame.

Volitionalists disagree amongst themselves about what exactly control amounts to: perhaps it requires reasons responsiveness, or sourcehood, or the ability to do otherwise. But they all agree that the moral appropriateness of blame requires control on the part of the agent being blamed. Anti-volitionalists deny this. Anti-volitionalists

---

86 Wolf (2011)
88 See Fischer and Ravizza, Derk Pereboom, and Peter van Inwagen respectively.
89 It can be someone difficult to clearly differentiate between volitionalism and anti-volitionalism since volitionalists do not all agree on what control amounts to. Thus, when explicating volitionalism generally we must rely on the imprecise notion of control.
normally seek to replace the control requirement for blame with something easier to meet. Angela Smith, for example, offers the following account:

**Rational Relations View**: An agent is, in principle, open to moral appraisal (including moral praise and blame) for something if and only if that thing reflects her rational judgment in a way that makes it appropriate, in principle, to ask her to defend or justify it.\(^9^0\)

We could take the Rational Relations View as an attempt to specify the particular kind of control needed for appropriate praise and blame. However, for our purposes it will be useful to reserve the term ‘control’ for the more robust requirements posited by the volitionalists. The Rational Relations account thus differs from standard volitionalism in that it allows for agents to be appropriately blamed for actions over which they did not exercise control, so long the actions adequately reflected their rational judgment.

One might worry that there is no genuine dispute here and that the anti-volitionalists are only talking about Scanlonian blame. However, anti-volitionalists such as Smith (2008) explicitly claim that their views apply to the reactive attitudes in particular, so we do have a genuine dispute between volitionalists and anti-volitionalists.

Now at first it might seem obvious that volitionalism is true. After all, it is common to offer and accept claims such as “he couldn’t help it” as reasons to cease blaming someone. And this seems to point to a control requirement for the appropriateness of blame. However, views like Smith’s can account for many cases in which these sort of claims excuse. For example, if someone failed to keep a promise to meet a friend because they were tied to a chair, Smith’s view could explain the inappropriateness of blame on the grounds that the failure to keep the promise did not

---

\(^9^0\) Smith(2008). Smith never states the view in precisely this way but this captures the substance of her view.
reflect the agent’s rational judgment. So we cannot conclude that volitionalism is true merely from the fact that we excuse agents in these types of cases. There are some cases, e.g. cases in which the agent’s values are “manipulated in”, in which only volitionalism appears to yield the correct result regarding the appropriateness of blame. However, anti-volitionalists have argued that there are also cases in which it is intuitively appropriate to blame agents for attitudes or behavior despite the fact that the agent appears to lack control. Thus, it is not entirely obvious that volitionalism should be preferred on merely intuitive grounds. In any case, my goal here is not weigh in concerning the intuitive evidence for or against volitionalism, but rather to solve a theoretical problem for volitionalism.

I: A Problem for Volitionalism

Volitionalism faces a significant problem. The problem is that it is difficult to see why blame must be morally inappropriate if the agent being blamed was not in control of her behavior. The problem I am interested in here is not the problem of accounting for why it would be wrong to treat someone who lacked control over her behavior harshly, say, by imprisoning her. I take it to be obvious that people deserve not to be treated in such ways and that such treatment would be seriously unfair. The problem I am interested in arises even if we grant that it is wrong to punish someone who lacks control over her behavior.

The problem arises because blame, unlike punishment, need not involve harsh treatment. In fact, you can blame someone without interacting with her at all. Given this

---

91 See Smith(2008) for an anti-volitionalist argument which appeals to these sorts of cases.
fact, what is it about blame that renders it morally problematic in cases where the agent lacks voluntary control? One potential answer is that blame somehow unfairly harms or burdens the agent and is thus wrong for the same reasons punishment would be wrong. However, Pamela Hieronymi has argued that there are serious problems with such views.92 Hieronymi maintains that the burdens imposed by blaming an agent need not be inappropriate even if the agent lacks voluntary control.

Consider the analogy between distrust and the reactive attitudes. It is does not seem to be morally inappropriate or unfair to distrust someone who lies repeatedly even if the person lacks volitional control. This is true despite the fact that your distrust imposes certain sorts of burdens on the person. Hieronymi admits that it may be in some sense unfair that the compulsive liar suffer the burdens of being distrusted. But it does not seem to follow from this that you have gone wrong by distrusting them.93 Fear provides another good example here. It can be entirely appropriate to fear a dangerous person even if the fact that she is dangerous is not under her control.

Hieronymi suggests that the burdensome force of distrust comes from the judgment that the person is untrustworthy. Given this, she claims that it is fair for you to distrust someone so long as they really are untrustworthy. This seems plausible in case of distrust. Now I do not think that it follows from the fact that a compulsive liar is untrustworthy that he deserves our distrust.94 Rather, I think that despite the fact that he does not deserve our distrust it is not unfair for us to distrust him. It should be clear that it

93 Hieronymi (2004) p. 119-120
94 Hieronymi (2004) appears to take the opposite view.
is not always unfair to harm or burden someone merely because they do not deserve it. If an employer fails to hire a qualified job applicant because he has more qualified applicants than open positions, he does not thereby treat the applicant unfairly, despite the fact that the candidate does not deserve to be passed over for the job. This sort of case contrasts with other cases, such as punishment, in which the agent who lacks control positively deserves not to be treated in a certain way. The compulsive liar, I claim, neither deserves to be distrusted nor deserves trust. Thus, it is not unfair to distrust him so long as he is untrustworthy.

Is something similar true concerning the reactive attitudes? It is controversial whether possessing emotions in general essentially involves making certain types of judgments. 95 But I am willing to grant to Hieronymi the claim that the reactive attitudes essentially involve making judgments. (Here and throughout I use the term ‘judgment’ as a catch-all for any attitude that has propositional content and aims at truth.) Hieronymi claims that the burdensome force of the reactive attitudes comes from the judgments that they carry with them. (In particular the judgment that the person being blamed has shown disregard or ill-will.) Thus, she claims, it is fair to blame someone just in case the judgments which the reactive attitudes carry with them are true. 96 It is difficult to see why this claim is less plausible than her parallel claim about distrust.

The problem for volitionalists, then, is to explain why blame is not like distrust. Volitionalists need to give an account of why blame, unlike other negative attitudes we

95 For theories of the emotions that endorse the idea that judgments are essential see Solomon (1980), Oakley (1992) and Nussbaum (2001). For a rejection of this view see Roberts (2003).
96 Hieronymi (2004).
might hold towards someone, is always inappropriate in the absence of control. To do this
volitionalists must maintain that blame differs from distrust because either (1) blame is
not appropriate whenever the judgments that form an essential part of the blaming
attitudes are true or (2) some judgments that form an essential part of the blaming
attitudes cannot be true unless the agent is in control of her behavior. Looking at a
(Hieronymi inspired) argument against volitionalism will help us to understand these two
potential ways a volitionalist might go about solving this problem.

Anti-Volitionalist Argument

P1. The Reactive attitudes are morally inappropriate only if some of the
judgments they essentially carry with them are false.
P2. None of the judgments which the reactive attitudes essentially carry with
them entail that the agent being blamed exercised volitional control.
C. So it can be morally appropriate to blame someone who did not exercise
volitional control.

An externalist volitionalist, let’s say, responds to this argument by denying P1.

Externalist volitionalists hold that there are some standards external to the content of the
reactive attitudes which renders them inappropriate despite the truth of their content.

Internalist volitionalists, by contrast, deny P2 and maintain that some judgment which
forms an essential part of the blaming attitudes cannot be true if the agent lacks volitional
control. In short it is the internal content of blaming attitudes which renders them
different from attitudes such as distrust and explains why volitionalism is true.

There is one way of attempting to undermine P1 which will be of no help the
volitionalist. It might be thought that P1 is undermined because whether the reactive
attitudes are appropriate depend not only on the facts about the person being blamed but
also on facts about the person doing the blaming. For example, facts about the nature of
the relationship between the agent blaming and the agent being blamed or facts about whether the blame would be hypocritical. These sorts of factors undermine the agent’s standing to blame. Hieronymi suggests that for each of these factors we will find some judgment in the reactive attitude itself which will be false when the agent lacks the standing to blame. However, even if this is false, it will be of little help to the volitionalist. The facts that call into effect one’s standing to blame don’t seem to be the right sort of facts to appeal to in explaining why control is needed for blame to be appropriate. These facts are about the person doing the blaming rather than about the one being blamed.

How plausible is externalist volitionism? I do not think it is hopeless. It might turn out that the reactive attitudes, unlike distrust, are governed by external standards that render them inappropriate even when the judgments that come along with them are true. However, I am not optimistic. The most obvious external standard to invoke is one that forbids adopting attitudes that dispose one to react negatively towards someone in response to actions beyond their control. However, this norm would apparently rule out distrust and fear as well. The externalist volitionalist might try to distinguish distrust from the reactive attitudes on the grounds that distrust of the untrustworthy generally has good consequences while the reactive attitudes do not. However, it is not at all clear that the reactive attitudes do not normally have good consequences. Guilt and knowledge of another’s resentment both often motivate one towards moral change, and could do so even in cases where the original bad behavior was involuntary.

---

97 Hieronymi (2004) p. 133
It is thus important for the prospects of volitionalism, I claim, that a plausible version of internalist volitionalism can be developed. This will be my aim in the next section two sections. I will try to identify some judgments essential to the reactive attitudes which entail that the agent being blamed had control over their behavior. I will begin by examining guilt and agent regret.

**II: Two Volitionalist Accounts of Guilt**

Agent-regret is a form of regret that occurs in response to things going badly due to the exercise of one’s agency. We can see that it is distinct from guilt because it is appropriate in cases where guilt is not. For example, someone who while driving their car in a completely non-negligent way hits and kills a child playing in the road should not feel guilt. However, it would be entirely appropriate to feel agent-regret. In fact, someone who did not feel agent-regret, someone who was not deeply pained by the fact that her actions caused a death, would be regarded as cold hearted.

It may seem quite difficult to distinguish between guilt and agent-regret. The affect, the painful feeling, experienced in both cases is not clearly distinct. Susan Wolf has suggested that we can distinguish guilt from agent regret by appealing to their content. I think this is the right approach. If there is any noticeable difference in the felt affect in each case I suspect that this difference is explained by the difference in the content of the attitudes. Wolf’s specific suggestion is that guilt is different from agent-regret because it “is at least partly a function of judging oneself to be blameworthy.”\(^9^8\)

This is insightful; judgments connected to the emotion of agent regret (absent guilt)

---

might include judgments like ‘I hurt her’ or ‘I’ve done something terrible’, but not the judgment that one is blameworthy.

Although I like Wolf’s general approach here, I do not want to accept it as is. If we analyze guilt in part by appealing to judgments of blameworthiness we appear to encounter a problematic circularity. This is due to the fact that Wolf’s account of what it is to judge that someone is blameworthy invokes the reactive attitudes:

… we may construe the judgment of blameworthiness as the judgment that a person is such as to deserve being the object of such an attitude [i.e. the reactive attitudes]... Anger, resentment, and the like are the prior concepts.  

I think something like this is right. But given this account it would be problematically circular to then define guilt or the other reactive attitudes by appealing to the notion of judgments of blameworthiness.

Another possible problem with appealing to judgments of blameworthiness to separate guilt from agent regret is that judgments of blameworthiness may be too narrow to do the job. Someone who has just hit a child with their car as a result of negligence may in their anguish be thinking that they deserve to die rather than judging that they would be appropriately resented. But of course they are still feeling guilt.

This second problem is not decisive. Wolf might reply that such a person would have the dispositional judgment that resentment is appropriate. But the case is suggestive of an alternative in the spirit Wolf’s suggestion which avoids the circularity problem as

---

99 Wolf (2011)

100 Sher (2006) points out a similar circularity.
well. I propose the following account (the first of my two suggested accounts for the volitionalist):

**Guilt/Desert Link:** A feels guilt only if A judges that he deserves something bad, and this is because judging that you deserve something bad is an essential part of what it is to feel guilt.\(^{101}\)

This account, like Wolf’s, uses a judgment that some negative reaction is appropriate to distinguish guilt from agent regret. It has the advantage of being broader than Wolf’s account: a judgment that one deserves anything from execution to having someone makes faces at you (or just resent you) will fulfill the requirement. **Guilt/Desert Link** also avoids circularity because it is plausible that desert, unlike blameworthiness, is not defined by appeal to the reactive attitudes.

One initial worry one might have with my account is that I am building too much propositional content into guilt. One can feel guilt, this complaint says, without judging much of anything at all. Feeling guilt is primarily a matter of affect.

I have three things to say about this. First, adopting an account of guilt on which it is almost completely affect leads us back to the problem of distinguishing guilt from agent regret. A defender of an affect based account would need to provide a way of distinguishing these two attitudes. I am not sure that this can be done without appealing to judgments. Second, the force of the objection to **Guilt/Desert Link** is greatly diminished if we allow for the judgment to be something like a dispositional belief. (Recall that I am counting as a judgment any attitude that has propositional content and aims at truth.) I agree that one can feel guilty without having an occurrent or conscious

\(^{101}\) Here and throughout, the notion of desert I have in mind is basic moral desert. I mean ‘judges that he deserves something bad’ to be inclusive of judgments about particulars bad things (‘he deserves to be punched in the face’) and general judgments (‘he deserves some bad thing or another’).
judgment that one deserves something bad. But it does not follow from this that no judgment at all is present. Third, if guilt did not essentially carry any judgments with it, then P1 of the Anti-Volitionalist Argument would clearly be false and externalist volitionism would be vindicated.

Another objection one might raise to Guilt/Desert Link is that it makes irrational guilt impossible. R. Jay Wallace points out that “one can feel guilt without believing that one has really done anything wrong.”¹⁰² I agree that irrational guilt is a genuine phenomenon. It is not, say, merely mislabeled agent regret. But since Guilt/Desert Link requires a judgment that one deserves something bad, a judgment that is at least intimately connected with the judgment that one has acted wrongly, how can I make sense of irrational guilt?

In my view agents who experience irrational guilt will possess inconsistent judgments. They will explicitly believe that they have done nothing wrong and do not deserve anything negative. But they will also always have some sort of judgment that they do deserve something bad. This claim is plausible because anyone experiencing irrational guilt will have at least some dispositions to behave as though they judged themselves to deserve something bad. One nice feature of the claim that these agents have inconsistent judgments is that it makes it easy to see why irrational guilt is irrational.

¹⁰² Wallace (1996) p. 40
I now want to examine one more objection. The objection is that positing **Guilt/Desert Link** is unnecessary to account for the distinction between guilt and agent regret. We can instead appeal to principles that anti-volitionalists are happy to accept. For example, proponents of Angela Smith’s Rational Relations view\(^\text{103}\) might find the following principle attractive:

**Guilt/Poor Rational Judgment Link**: A feels guilt for X only if A judges that X reflects poorly on his rational judgment (in a way that makes it appropriate, in principle, to ask him to defend or justify it) and this is because judging that X reflects poorly on your rational judgment is an essential part of what it is to feel guilt for X.

Other anti-volitionalists would prefer slightly different accounts. A proponent of Hieronymi’s view might endorse something like:

**Guilt/Ill Will Link**: A feels guilt only if A judges that he has shown ill will, and this is because judging that you have shown ill will is an essential part of what it is to feel guilt.\(^\text{104}\)

These principles flow naturally out of anti-volitionalist theories of blame. Furthermore, they appear to provide an explanation of the difference between guilt and agent regret. The person who feels mere agent regret does not judge that they have shown ill will or poor rational judgment. The person who feels genuine guilt does judge that they have shown ill will or poor rational judgment. Thus, the anti-volitionalist might insist that we have no need of **Guilt/Desert Link**, principles that they already endorse will do just as well.

Even if I could not show that **Guilt/Desert Link** is any more plausible than the anti-volitionalist alternatives, I have still identified a plausible principle which (I will

---

\(^{103}\) See Smith (2008).

\(^{104}\) Hieronymi (2004) appears to hold that the judgment that a person has shown ill will or disregard is an essential component of the negative reactive attitudes.
argue) volitionists can appeal to in order to deny P2 of the Anti-Volitionist Argument. Having said this, I also doubt that Guilt/Poor Rational Judgment Link or Guilt/Ill Will Link can provide anti-volitionists with the needed distinction between guilt and agent regret.

Suppose that Bill has been prescribed a painkiller to take in the aftermath of a medical procedure he has recently undergone. Suppose also that this particular painkiller has the heretofore unidentified side effect of causing those who take it to weigh moral reasons for action more lightly than they ought. The drug does not bypass deliberation; rather it causes a different weighing of reasons in deliberation. We can even suppose that it does this by starting a causal chain which is subsequently identical to standard causal chains which result in a change in how an agent weighs reasons. As a result of the drug, Bill assigns less weight than he ought to his reasons to be loyal to his partner, Alice, and he proceeds to cheat on her.

Now anti-volitionists will probably want to say that on at least some of the ways in which we could spell out the details of how the drug works Bill is blameworthy for his behavior. At least some volitionists, on the other hand, will deny that Bill is blameworthy. Setting this issue aside for the moment, I want to explore the various ways in which Bill himself might coherently react to his behavior.

Let us suppose that Bill is a volitionist. Furthermore, when Bill learns how the drug effected his decision making process, he concludes that he did not exercise

---

105 Especially since the drug need not operate in a manner relevantly different from how standard causal processes operate on an agent’s will.

106 Especially if the causal chain from taking the drug to cheating on Alice is deterministic.
volitional control over his behavior. And thus, that feeling guilt would be inappropriate.

Given all this, it seems that Bill might coherently feel agent regret, but not guilt, for his behavior. Let us add that Bill recognizes that he has shown ill will towards Alice and that his behavior reflects poorly on his rational judgment. Might he not still feel agent regret, but not guilt? In fact it seems to me that he might feel agent regret (but not guilt) about the very fact that he had shown ill will. Indeed, I suggest, his emotion of agent regret might itself have the content ‘I showed Alice ill will’, and yet he still need not be feeling guilt. Given his belief that his behavior does not render guilt appropriate in this case this all seems perfectly natural.

I see two options available to the anti-volitionalist here. One would be to dig their heels in and claim that Bill is feeling guilt. This seems to me to get the case entirely wrong. Bill is intuitively feeling agent regret, not guilt. Something that is required for guilt still seems to be missing. The point generalizes to any agent who does not believe that they are responsible for or deserving of a negative response for their ill will. Consider someone who believes that determinism is true and that given this he is not responsible for his behavior. It makes perfect sense for such a person to feel agent regret, and not guilt, in response to his displays of ill will.

The second move the anti-volitionalist might make is to say that Bill is not displaying ill will (or judging that he displays ill will) in the sense that they have in mind. Ill will is displayed and rational judgment is poorly reflected, only when the agent is not unduly influenced by outside factors such as Bill’s painkiller. Now this might vindicate Guilt/Poor Rational Judgment Link and Guilt/Ill Will Link as plausible accounts of
the difference between agent regret and blame but only at a terrible cost to the anti-
volitionalist. By making this move the anti-volitionalist would be granting that there is a
control condition on the appropriateness of guilt. What motivates volitionalism is the
thought that certain forms of undue influence (e.g. some types of manipulation of an
agent’s reasons) undermine the agent’s control and thus render blame inappropriate. The
anti-volitionalist who offers this response would be accepting the volitionalist claim and
just building the control condition into the concept of showing ill will or poorly reflecting
one’s rational judgment.

The preceding point motivates my second suggested volitionalist friendly account.
Let’s make a distinction between two notions of ill will:

- **Weak Ill Will**: an attitude that can be displayed in the absence of agential control.
  (A sort of disregard that need not be a product of free agency)

- **Strong Ill Will**: an attitude that can be displayed only if agential control is
  present. (Disregard that is a product of free agency)

And here is my second suggested account:

- **Guilt/Strong Ill Will Link**: A feels guilt only if A judges that he has shown strong ill
  will, and this is because judging that you have shown strong ill will is an essential part of
  what it is to feel guilt.

  **Guilt/Strong Ill Will Link** is motivated because it can account for Bill’s not
feeling guilt while a **Guilt/Weak Ill Will Link** principle could not. One might worry that
Strong Ill Will is a gerrymandered notion. But this is not obvious. It is plausible that there
is a pre-theoretical agency laden notion of ill will which is captured by Strong Ill Will
(‘will’ is in the name after all).
In my view both Guilt/Desert Link and Guilt/Strong Ill Will Link are intuitively plausible. They are consistent with each other so we need not choose between them. Furthermore, endorsing either of them provides a nice account of the difference between guilt and agent regret. I will now discuss whether they can be expanded so as to apply to other blaming attitudes and whether the expanded versions enable the volitionalist to reject P2.

III: Expanding the Account

I now want to propose that analogues of Guilt/Desert Link and Guilt/Strong Ill Will Link are plausible with regard to resentment and indignation as well (and any other reactive attitudes that are instances of blame). Here are the more general versions I find plausible:

**Blame/Desert Link:** A blames B only if A judges that B deserves something bad, and this is because judging that B deserves something bad is an essential part of what it is to blame B.

**Blame/Strong Ill Will Link:** A blames B only if A judges that B has shown strong ill will, and this is because judging that B has shown strong ill will is an essential part of what it is to blame B.

There are at least three considerations in favor of endorsing these more general principles.

First, resentment and indignation and the judgment that the agent deserves something bad tend to go together. Normally if you resent someone you would also be inclined to assert that they deserve a negative response. Additionally and suggestively, resentment tends to dissipate if one comes to the view that the agent does not deserve a
negative reaction.\(^{107}\) (Note that there is no such tendency in the case of distrust.) This tendency for resentment to dissipate is plausibly accounted for by positing that there is rational pressure against feeling resentment towards someone while judging that they deserve nothing bad. And this rational pressure is easy to make sense of if judging that X deserves something bad is an essential part of what it is to feel resentment towards X.

This sort of consideration supports **Blame/Strong Ill Will Link** as well. There is also rational pressure to judge that those you resent have shown ill will. However, a **Blame/Weak Ill Will Link** type principle could account for this as well. So, in the case of **Blame/Strong Ill Will Link**, the point is not decisive.

Second, although not as clear as in the case of guilt and agent regret, it does seem that there are angry attitudes which we can feel towards or about someone’s behavior which stop short of genuine resentment or blame. Consider Bill’s case from Alice’s perspective. Suppose that Alice is also a volitionalist and also judges that blaming Bill would be inappropriate. It seems that she might nonetheless rationally be angry that Bill cheated on her, that his agency was caught up in this betrayal of his commitment to her. In short, she might feel a second personal analogue of agent regret. If this is right, then endorsing **Blame/Desert Link** and/or **Blame/Strong Ill Will Link** can allow us to distinguish the second personal analogue of agent regret from genuine resentment.

Third, it would be implausible to hold that feeling guilt requires a judgment of desert or ill will while feeling resentment and indignation do not. This would amount to positing a significant asymmetry between the appropriateness conditions for self-blame

\(^{107}\) I explain exceptions to this by positing inconsistent judgments.
and the appropriateness conditions for blame by others. We would then have many cases in which it would be entirely appropriate for others to blame an agent, but inappropriate for the agent to feel guilty. But this is implausible. Thus the evidence in favor of guilt essentially requiring the judgment that the agent showed ill will or deserves something bad is also evidence that there is a similar requirement for the other blaming attitudes.

Now Recall P2 of the Anti-Volitionalist Argument:

P2. None of the judgments which the reactive attitudes essentially carry with them entail that the agent being blamed exercised volitional control.

**Blame/Strong Ill Will Link** clearly undermines the P2. But is **Blame/Desert Link** of any help to the volitionalist? In order for it to aid the volitionalist in rejecting P2, there must be a principle connecting desert and volitional control. The principle I favor is the following:

**Desert/Control Link:** B deserves something bad only if B has exercised his free agency in a morally bad way.

Given **Blame/Desert Link** and **Desert/Control Link** P2 is false. All of the blaming reactive attitudes carry with them the judgment that the agent deserves something bad and this judgment cannot be true unless the agent exercised volitional control. **Desert/Control Link** is notably silent on what the connection must be between the agent’s exercise of free agency and the particular act, attitude, character trait, etc. for which the agent deserves a negative response. These issues are quite complex and I cannot discuss them here. But the very general connection between free agency and desert posited by **Desert/Control Link** will be sufficient for my purposes. Furthermore, I give no analysis of free agency; this is because I want my account to neutral between the
various volitionalist theories of free agency. All I will say about the nature of free agency is that it involves substantial control of the sort volitionalists think is required in order for blame to be appropriate.

An anti-volitionalist presented with Desert/Control Link might be tempted to complain that I have made little progress. We began with the problem of accounting for why attitudes such as guilt and resentment are morally inappropriate unless the agent exercised control. All we have done, the complaint goes, is move to the problem of accounting for why an agent cannot deserve something bad unless he has exercised control. This problem, however, does not pose the same challenges as the original one. This is true in part because there does not appear to be a close analogy between moral desert and some other property an agent can have without exercising control. Thus we do not encounter anything parallel to the issue of explaining why resentment differs from distrust.

Furthermore, most volitionalists will be inclined to find Desert/Control Link highly plausible. It captures the volitionalist friendly intuition that agents who commit serious crimes but lack control over their behavior do not deserve to be punished, while criminals who did possess control over their behavior do deserve to be punished. Since I am trying to solve a problem for volitionalism, rather than attempting to persuade anti-volitionalists to become volitionalists, it is acceptable to appeal to tools in the volitionalist tool box to solve the problem.

At this point it appears that volitionalists can comfortably reply to the Anti-Volitionalist Argument by endorsing Blame/Desert Link and Desert/Control Link and
defending them from objections. Since most volitionalists would be inclined to accept Desert/Control Link anyway, we have provided them with a response that does not require that they take on any new commitments (beyond Blame/Desert Link). Of course anti-volitionalists might attack Desert/Control Link by presenting cases in which it seems that agents who lack control over their behavior nonetheless deserve a negative response. But this debate is far removed from the special problem for volitionalism posed by the Anti-Volitionalist Argument. This is just a return to the debate over whether volitionalists or anti-volitionalists better capture our intuitions about cases. My goal here is not to settle that debate in favor of volitionalism. Rather I hope to have shown that, given the acceptance of a volitionalist framework, a plausible explanation of why the reactive attitudes are inappropriate absent control can be given. Volitionalism may turn out to be false, but not because it lacks an account of why the blaming attitudes are inappropriate in cases where attitudes such as fear and distrust are perfectly fine.

It is important to see that I have not defended volitionalism by arguing that agents who do not exercise volitional control deserve for us to refrain from resenting them. It may be that, just as compulsive liars do not deserve trust or distrust, so too compulsive wrongdoers deserve neither resentment nor the lack of resentment. My defense of volitionalism claims that some of the judgments that the reactive attitudes essentially carry with them are false if the agent lacks control. I have not claimed that agents who do not have control over their actions deserve for us to refrain from resenting them.

IV: Some Objections
I now want to consider some objections that Hieronymi has raised to accounts similar to those I have developed. She briefly discusses a view that builds something like judgments about desert into the content of resentment:

One might, for example, claim that [the reactive attitudes] are sensitive not only to the judgment that ill will has been shown, but also, crucially, to the judgment that, because of this ill will, one is entitled to treat the wrongdoer in certain ways…This account would allow blame both to carry a certain force and to be rendered unfair by whichever conditions render such treatment unfair.\(^{108}\)

This account, like **Blame/Desert Link**, would indirectly make blame sensitive to the same sorts of considerations that render punishment unfair. Hieronymi dismisses it on the grounds that “There is more to the force of blame or the force of the reactive attitudes than either affect or the threat of or the entitlement to certain forms of treatment.”\(^{109}\)

I agree that the judgment that the agent deserves something bad does not exhaust the force or significance of the reactive attitudes. Much of this force is probably accounted for by the judgment (favored by Hieronymi) that the agent has shown disregard or ill will (of either the weak or strong variety). However, I do think it is plausible that some of the force of the reactive attitudes is explained by the judgment that the agent deserves something bad. Consider Bill and Alice again, suppose that, instead of merely feeling the second personal analogue of agent regret, Alice begins to full on resent Bill. This resentment will have a different burdensome force than her previous attitude. It will create a further issue that Bill needs to appropriately respond to if the damage to their relationship is to be repaired. For example, Alice’s resentment plausibly calls for more

\(^{108}\) Hieronymi p.122

\(^{109}\) Ibid.
apologetic behavior on Bill’s behalf. Now if the difference between Alice’s attitudes is accounted for by the new judgment that Bill deserves something bad, then the difference in force must also be explained by this judgment. In addition, the judgment that Bill deserves something bad plausibly gives Alice reason to wait longer before, and possibly to completely forgo, resuming her relationship with Bill. This clearly increases the burdensome force involved in resentment. Thus, I claim that the judgment that the agent deserves something bad can importantly contribute to the force of resentment and the other reactive attitudes, even if it does not exhaust it.

Hieronymi also considers views which build judgments about control directly into the reactive attitudes. She grants that there may be attitudes that contain these sorts of judgments, but she suggests that the reactive attitudes identified by Strawson do not contain such judgments.110 Strawsonian reactive attitudes are “non-detached” in a way that makes it implausible that judgments about control are part of their content. Hieronymi writes:

[The reactive attitudes] are thus sensitive to just those facts that make the wrongdoing interpersonally important—facts about one’s own standing, the standing of the victim, the standing of the wrongdoer, and the seriousness of the disregard. They are not sensitive to facts that do not change, in some understandable way, the significance of the wrongdoing for one’s interpersonal relations.111

Hieronymi’s view is that this constraint on the content of the reactive attitudes will rule out judgments such as “the person could have done otherwise” and other judgments about the agent’s control from being included in the content of the reactive attitudes. However

110 See Strawson (1962)

111 Hieronymi p. 135
her constraint may fair against views that include judgments directly about control in the reactive attitudes, I do not think it succeeds in ruling out **Blame/Desert Link.** This is because the fact that the agent deserves something bad in virtue of his wrongdoing does alter the significance of his wrongdoing for interpersonal relations. As we have already seen, it can change how others ought to respond to the agent.

Let me give another example of this. Consider how you should respond to a friend of yours who has caused the death of a child under circumstances which you regard as completely excusing. You thus judge that he does not deserve anything bad. It seems natural that you would (and should) respond to him by offering him comfort and perhaps by offering encouragement to modify his behavior in order to avoid similar accidents in the future. If, on the other hand, you judge that he does deserve a strong negative response, your behavior towards him will (and should) change in a quite different way. You will probably treat him very coldly and withdraw from various relationships you had with him. Thus your friend’s wrongdoing has greatly different implications for interpersonal relations when he deserves a negative response than it does when he does not deserve such a response.

One might worry that Hieronymi’s objection has more force against **Blame/Strong Ill Will Link.** However, it is not clear to me that this is right. It seems plausible that “seriousness of the disregard” is affected by whether or not the ill will displayed by the wrongdoer was a product of free agency. But a volitionalist who does not find this plausible may have reason to endorse **Blame/Desert Link** instead.
To sum up, I take myself to have shown is that there are two plausible accounts of the content of the reactive attitudes (the accounts captured by Blame/Desert Link and Blame/Strong Ill Will Link) that provide plausible replies to the Anti-Volitionalist Argument. These accounts also show why the reactive attitudes differ from attitudes such as distrust which can be appropriate even if the agent lacked control. If volitionalism is false, it is not because it cannot explain why the reactive attitudes are only appropriate when the agent exercised control.
Chapter 5: Ability, Foreknowledge and Explanatory Dependence

Many philosophers wish to maintain that the ability to do otherwise is compatible with comprehensive divine foreknowledge but incompatible with the truth of causal determinism. One problem with this view is that the Fixity of the Past principle which underlies the rejection of compatibilism about the ability to do otherwise and determinism appears to generate an argument for the incompatibility of the ability to do otherwise and divine foreknowledge as well. By developing an account of ability which appeals to the notion of *explanatory dependence*, we can replace the Fixity of the Past with a new principle which does not generate this difficulty.

I: The Problem

Many philosophers are attracted to both of the following views:

(1) The ability to do otherwise is compatible with comprehensive divine foreknowledge.

(2) The ability to do otherwise is incompatible with causal determinism.

I will call the conjunction of (1) and (2) COMBO. There appears to be a significant tension internal to COMBO, however. This is because the very considerations usually appealed to in order to support (2) also appear to undermine (1). In particular, explanations of the truth of (2) usually appeal to something like the following principle:

Fixity of the Past (FP): An agent S can (at time \( t \) in world \( w \)) do \( X \) at \( t \) only if there is a possible world \( w^* \) with the same past up to \( t \) in which S does \( X \) at \( t \).
The obtaining of causal determinism guarantees that facts about the initial state of the universe together with facts about the laws of nature entail every fact about the future. So FP, together with a similar Fixity of the Laws principle, can be used to generate an argument for the incompatibility of the ability to do otherwise and causal determinism.

The problem is that FP appears to generate an argument for the incompatibility of the ability to do otherwise and divine foreknowledge as well. Since God is supposed to be infallible, there is no world where he falsely believes that you perform a certain act. And since his beliefs (if they really constitute foreknowledge) are part of the past, there seems to be no world with the same past in which you do otherwise than what God believed you would do. Thus, given FP, you cannot do otherwise. So COMBO’s problem is that FP is needed to provide an argument for (2), but FP also generates an argument against (1).

II: Possible Solutions

The traditional Ockhamist solution to this problem is to distinguish between the ‘hard’ and ‘soft’ facts about the past, where soft facts about the past are in some sense temporally relational, and thus, not strictly facts about the past. For example the fact that Kennedy was shot is a hard fact about the past. But the fact that Kennedy was shot 52 years before I wrote this paper is a soft fact about the past. With this distinction at hand, it is then claimed that facts involving God’s beliefs about the future are soft facts and that only the hard facts about the past must be fixed.

---

112 Some philosophers simply define causal determinism in terms of this entailment.
One problem with this solution is that we lack a good account of the difference between hard and soft facts that clearly places God’s beliefs in the soft category. In addition, there are plausible arguments for the conclusion that God’s beliefs are hard facts about the past (or at least contain what Fischer has called “hard elements”). At any rate, in what follows I will set Ockhamism aside and explore a different solution.

The prospect of a distinct solution is now emerging. Trenton Merricks has recently defended (1) on the grounds that “God’s beliefs about what an agent will do in the future depend on what that agent will do in the future.” And Michael Bergmann has suggested that the crucial difference between God’s beliefs and causal determinism is that God’s past beliefs are held “because of what I’m doing now, not vice versa.” Of course for this to be a solution to COMBO’s problem we would need an explanation of why these claims about dependence matter for an agent’s abilities. But what these proposals suggest is that the solution to COMBO’s problem is not to be found in the distinction between hard and soft facts, but rather by exploring the relationship between dependence and ability.

Let’s call a potential solution to COMBO’s problem that relies on the notion of dependence a DEPENDENCE SOLUTION. Of course, there is a natural way of interpreting Ockhamism on which it is a type of DEPENDENCE SOLUTION.

---

113 See e.g. Fischer (1994) and Todd (forthcoming).

114 Merricks 2011 p.567

115 Bergmann’s correspondence with Fischer, as reported in Fischer 2011 p. 222.

116 Other recent proponents of something like a DEPENDENCE SOLUTION include McCall (2011) and Westphal (2011).
Ockhamism relies on the view that at least some soft facts depend on our choices but the hard facts do not. However, I am here interested only in non-Ockhamist DEPENDENCE SOLUTIONs. That is, I am only interested in solutions that eschew worrying about whether particular facts are hard or soft and instead talk directly about dependence. (More on this distinction in section VIII.) Now, I suggest that an adequate DEPENDENCE SOLUTION should do two things:

(a) Identify the relevant sort of dependence

(b) Provide an account of ability which appeals to dependence and which renders COMBO plausible.

In what follows I will attempt to develop a DEPENDENCE SOLUTION that meets both of these criteria.

III: What Sort of Dependence?

In order to vindicate COMBO, we need our account of ability to appeal to a sort of dependence which plausibly yields a significant asymmetry between God’s past beliefs and determinism. Not just any sort of dependence will provide this result. In his discussion of Bergmann’s suggestion John Martin Fischer argues that counterfactual dependence in particular is not well suited to play this role:

On some views of the relevant counterfactuals, if causal determinism is true and I actually perform some action X, the following ‘backtracker’ is true: ‘If I were to refrain from X, the past would have been different all the way back’. Thus, on this sort of view of counterfactuals, there would indeed be a counterfactual
dependence of the past causal facts on the behavior in question, so it would not be obvious that the relevant notion of ‘because of’ would be asymmetric.\footnote{Fischer 2011 p. 223}

I think this line of reasoning is correct. It is plausible that facts about the initial state of the universe counterfactually depend on our choices. So appealing to the fact that God’s past beliefs also counterfactually depend on our choices is not a promising way to defend COMBO. The other option would be the view that the laws of nature (rather than the initial state of the universe) counterfactually depend on our choices. This option renders counterfactual dependence equally unsuitable for use in a defense of COMBO.

Fortunately for defenders of COMBO, there is a more promising notion of dependence available. I suggest that the DEPENDENCE SOLUTION appeal to the notion of \textit{explanatory dependence}. Although explanation has sometimes been thought to be a pragmatic notion, the recent literature on dependence and grounding has found room for an objective notion of explanation. This provides an opportunity to develop the DEPENDENCE SOLUTION in a plausible way. Like necessity, explanatory dependence can come in different varieties. Just as we have, logical, conceptual, metaphysical, and natural necessity, we also have corresponding types of explanatory dependence. Here are some examples from the literature (where ‘A because B’ means ‘A explanatorily depends on B’):

- Logical: Sam is ill or 2+2=5 because Sam is ill.
- Conceptual: The vase is colored because it is red.
- Metaphysical: The set \{Socrates\} exists because Socrates does.
Natural: Sam died because John stabbed him in the heart.  

Explanatory dependence is the broad notion that captures what all of these different cases of dependence (and perhaps other types of cases as well) have in common. This is the notion of dependence I will appeal to in developing my DEPENDENCE SOLUTION.  

IV: A Partial Account of ‘Can’

So far I have been using the term ‘ability’. However, talk of abilities is ambiguous between general abilities and “in the moment” abilities. This is because the later notion involves opportunities as well as general abilities. The following case illustrates the distinction. Suppose Bob is an excellent piano player who is currently tied to a chair far from any pianos. In this case Bob has the general ability to play the piano, but lacks the opportunity to play the piano. Following Christopher Franklin (2011), we can use the term ‘can’ so that it applies to an agent when and only when he has both the general ability and the opportunity to perform the act in question. Here is Franklin’s “semi-stipulative” way of understanding ‘can’:

An agent S can φ at t in possible world W iff S has the [general] ability and opportunity to φ at t in W.  

---

118 These examples come from Correia (2005) and (2008) and Schnieder (2006).

119 Philosophers who reject the idea that there is an objective notion of ‘explains’ or ‘because’ may still be able to adopt an account similar to mine. My view is that some ‘because’ claims capture the objective structure of the world and as a result we can use the terminology of ‘because’ claims to develop an objective account of ability. So long as one accepts that there is an objective structure of the world, it should be possible to develop an account similar to mine using whatever terminology does capture this structure.

120 p. 695. Franklin’s formulation follows Austin (1956).
I shall presume that it is this sense of ‘can’ that defenders of COMBO are concerned with. They want to show that an agent can have both the general ability and opportunity to do otherwise despite God’s foreknowing what he will do, but that an agent cannot have both the ability and opportunity to do otherwise if his act is causally determined.

With this framework available, we can now give a partial account of ‘can’ which appeals to explanatory dependence. I will not attempt to give a precise account of general abilities. (They appear to involve having the disposition to succeed in performing the act in a wide array of circumstances.) And I think it should be granted that general abilities are compatible with both determinism and divine foreknowledge. Rather I think it is promising to use our notion of dependence in order to analyze the opportunity condition. Here is my suggested account:

**No Independence Account**: S has the opportunity to do A at T in W iff there is a possible world in which all of the facts in W that do not explanatorily depend on X’s choice at T still obtain and X does A.

The idea behind this account is that I do not have the opportunity to perform an act if facts which are explanatorily independent of my choice entail that I do not perform the act. And I take it that this suggestion will appear intuitively plausible to those inclined towards incompatibilism about determinism and the ability to do otherwise.

**V: Using the Account to Save COMBO**

The **No Independence Account** leads to the following principle:
Fixity of the Independent Past (FIP): An agent $S$ can (at time $t$ in world $w$) do $X$ at $t$ only if there is a possible world $w^*$ in which all of the facts in $w$ up to $t$ which do not explanatorily depend on $S$’s choice hold and $S$ does $X$ at $t$.

In my view the DEPENDENCE SOLUTION can vindicate COMBO by endorsing the Fixity of the Independent Past and rejecting the original Fixity of the Past (FP) principle.

If we replace the original FP principle with FIP we can avoid the problem the Fixity of the Past created for compatibilism about foreknowledge and the ability to do otherwise. So long as God’s beliefs explanatorily depend on our future choices, FIP does not yield the result that we cannot do otherwise than God believes we will do.

Furthermore, this dependence claim is plausible if we assume that causal determinism is false. Given the falsity of determinism, it appears that there are no present or past conditions sufficient to guarantee that certain future choices will be made. But, given God’s essential omniscience, it would appear that the full explanation of God’s beliefs must include something that does guarantee the truth of his beliefs. Thus, his beliefs must be explained by something located in the future, most plausibly by the choice itself.

Trenton Merricks (2009) has also suggested an argument which (when modified a bit) supports the conclusion that God’s beliefs depend on our future choices. The argument runs along the following lines: It is true that Jones sits at $t$ because of Jones’s act of sitting at $t$. Furthermore, God believes that Jones sits at $t$ because it is true that Jones sits at $t$. (This second claim is quite plausible given God’s essential omniscience).

---

121 Merrick’s original claim is “that Jones sits at $t$ was true a thousand years ago because Jones will sit at $t$”(p.54). My version of the argument makes it clearer that Jones’s act itself is doing the explaining.
So, assuming that the relevant notion of dependence is transitive, God’s belief does depend on Jones’s act.

So we have good reason to think the FIP will not generate the same problem for compatibilism about freedom and foreknowledge that FP did. Furthermore, FIP can still be used to generate an argument against compatibilism about freedom and determinism. This is because it is not plausible that the initial conditions of the universe depend on our choices. Rather, it is plausible that our choices depend (at least partially) on the initial conditions of the universe. So FIP and the No Independence Account can vindicate COMBO’s claims concerning freedom and determinism. (We also need the claim that the laws of nature do not depend on our choices, which seemingly involves rejecting the Humean conception of laws.)

VI: Why FIP Rather than FP?

So we have seen that it is useful for defenders of COMBO to replace FP with FIP. But is this enough to justify rejecting FP? After all, it is intuitive that the past is fixed. Furthermore, Fischer (2011) has argued that FP can be motivated by the claim that hard facts about the past are “over-and-done-with”. And if FP is true, then even if God’s beliefs explanatorily depend on our future choices, we still have not solved COMBO’s problem. In short, those who are inclined to accept the DEPENDENCE SOLUTION need to deal with the plausibility of FP. Fischer and Tognazzini (forthcoming) press the worry this way:

But how exactly does the dependence point in any way vitiate – or even address – the point about the fixity of the past? That is, if a hard fact about the past is now
fixed and out of our control precisely because it is “over-and-done-with”, why is the dependence in question relevant? If fixity stems from over-and-done-with-ness, and over-and-done-with-ness is a function of temporal intrinsicality, both of which seem plausible, then it would seem more reasonable to conclude that even the dependent hard facts are fixed. (p. 22)

My view is that FP and the “over-and-done-with-ness” claim (insofar as it is interpreted to support FP) are both intuitive because we tend to assume that the hard past is explanatorily independent of future events. Once we drop that assumption, the intuitive plausibility of FP is greatly reduced.

Imagine you have come to believe that you are sitting in a working time machine. (Set aside the issue of whether time travel is genuinely metaphysically possible.) You believe that the machine is programmed such that if you push the big green button in front of you, you will travel back 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 green button. Note that by accepting the possibility of time travel you have dropped the assumption that the past must be explanatorily independent of the future.

I claim that in this circumstance FP would no longer be an intuitive principle. If you accept FP, then you should accept that either you cannot push the button or you cannot refrain from pushing the button. After all, it is either a fact about the past that you appeared in 1492 or it is a fact that you did not. And you believe that there is no world with the same past and laws in which you push the button and do not travel through time, or vice versa. (Here I assume that you accept the fixity of the laws principle.) However, I
do not think that this claim about your lack of options would seem true to you. Surely it would seem that you have the option to push the button and the option to refrain from pushing the button. It would not seem that the past was “over-and-done-with” in any sense inconsistent with your freedom.

Of course such intuitions are not infallible and I have not established that FP is false. But I do think that reflection on such cases reveals that what may look like intuitive support for FP is really only support for FIP. Once we drop the assumption of independence, FP is not so intuitive.

One might worry that intuitions concerning ability and time travel are unreliable in general. Consider the notorious case in which Bob travels back in time with the intent of killing his own grandfather (prior to the conception of his father or mother). If Bob finds himself standing in front of his grandfather holding a loaded gun, it will surely seem to him that he could kill the man. However, it is not clear that Bob’s seeming is correct. Since there is no possible world where the man is really is Bob’s grandfather and Bob causes him to permanently cease to exist prior to the conception of Bob’s father or mother, I am inclined to think that Bob cannot kill the man in front of him.\(^{122}\)

Should consideration of cases like this lead us to be skeptical of intuitions concerning ability and time travel in general? I don’t think so. We are tempted to reject Bob’s intuition that he can kill his grandfather because accepting it generates special difficulties. The fact that some intuitions in a domain generate difficulties is not normally a good reason to be skeptical of all intuitions within that domain. For example, we should

\(^{122}\) Of course, not all philosophers agree with this. See e.g. Lewis (1976)
not reject all intuitions about value because some intuitions about value lead to the
Repugnant Conclusion.

I now want to offer two more considerations which, in my view, count against the
view that FP is true even if the past depends on the future. First, it seems to me that
something that depends on my choice is not the sort of thing that can limit my options in
making the choice. If FP holds and facts about the past (e.g. God’s beliefs) depend on my
choices, then these dependent past facts constrain my behavior. But things that depend
on my choices are derivative; they are not the sort of things that can constrain my choice.

Second, and relatedly, this sort of view risks endorsing explanatory circles. In
many cases, if something limits my options it thereby (partially) explains why I choose as
I do. But if we say that my choices explain God’s beliefs and that God’s beliefs explain
my choices we are left with a rather unpleasant circularity. In order to avoid this,
defenders of FP would need to provide a plausible account of how God’s beliefs
constrain my behavior without explaining my behavior. It is not clear to me that such an
account can be given.

Of course, cases involving counterfactual interveners (such as the Frankfurt cases)
show that it is possible to limit an agent’s options without actually explaining his
behavior. However, I do not think that the Frankfurt cases provide a good model for
Divine Foreknowledge. The counterfactual intervener limits opportunities because he
stands ready to intervene if necessary. But it is not clear why God (or God’s
foreknowledge) should be conceived as standing ready to prevent an agent from acting in

---

123 See Frankfurt (1969)
a manner contrary to God’s beliefs. Furthermore, since counterfactual interveners do explain the agent’s behavior in some other possible world, claiming that God’s foreknowledge functions like a counterfactual intervener appears to involve positing an explanatory circle in the world where God’s foreknowledge does explain the agent’s choice. So I do not think that appealing to cases involving counterfactual interveners will avoid this objection.

Cases involving counterfactual interveners also show that my two points are not independent of each other. If explanatory circles are possible, then it will probably be possible for facts that depend on my choice to nonetheless constrain my behavior. (Imagine traveling back in time and becoming your own counterfactual intervener.)¹²⁴ In my view this gives us a further reason to reject the possibility of explanatory circles. It shouldn’t turn out that derivative facts can constrain my choices.

Overall I conclude that if the past can depend on the future, FP should be rejected. Thus, I think that the prospects for the DEPENDENCE SOLUTION as I have developed it look promising indeed.

VII: the DEPENDENCE SOLUTION and providential control

My DEPENDENCE SOLUTION, unlike Ockhamism, allows for the possibility that we can perform an act even if there is no world with the same ‘hard’ past in which we perform the act. This increases the amount of providential control God could exercise consistent with human freedom.

Consider the following case offered by Alvin Plantinga:

¹²⁴ I owe both the point and the example to Chris Tweedt
“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.”

Clearly it is providentially advantageous for God if this story is consistent with Paul having the ability to mow his lawn. God would then have the ability to ensure more outcomes while still giving us the ability to do otherwise. However, it is implausible for Ockhamists to maintain that Paul can do otherwise. In order to get the result that Paul can do otherwise, Ockhamists would have to say that one of the following is a soft fact:

(a) God intended to keep the ants away from lawns that would be mowed.

(b) The ants were in the yard.

But neither of these are plausible candidates for being soft facts. (b) is a paradigm case of a hard fact. And even if Ockhamists can meet the challenge of giving us an account on which it is plausible that God’s beliefs are soft because of the manner in which they are related to the future. It is a further challenge to defend the claim that God’s intentions

125 Plantinga 1986 p. 254
about how to act in response to his beliefs about the future are not hard facts about the past. It seems clear to me, insofar as I have a grip on the notion of a hard fact, that facts like (a) are hard facts about the past.

On the DEPENDENCE SOLUTION, however, the above story is consistent with Paul’s being free to mow his yard. This is because the ants being in the yard is a hard fact that is explanatorily dependent on Paul’s choice. So I am inclined to think that the DEPENDENCE SOLUTION increases providential control by allowing God more options for controlling outcomes while not undermining our freedom.

VII: the Uniqueness of the DEPENDENCE SOLUTION

My claim that the DEPENDENCE SOLUTION differs from Ockhamism in its implications for Divine Providence suggests that the DEPENDENCE SOLUTION is a unique solution to COMBO’s problem. Fischer and Tognazzini raise the worry that the DEPENDENCE SOLUTION is not unique. They offer the following comment on the DEPENDENCE SOLUTION:

In fact, we’re inclined to think that it just is the Ockhamist response, at least on a certain understanding of Ockhamism.

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 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, [DEPENDENCE SOLUTION proponents like] Merricks, McCall, and Westphal (on our charitable revision of their arguments) are just articulating a version of Ockhamism. (P. 20-21)

In my view, this gets Ockhamism, or at least the standard contemporary articulation of Ockhamism, wrong. Ockhamism should be construed as being concerned with both temporal intrinsicality and dependence. Ockhamism provides a method for determining which facts could depend on the future. It says that only temporally relational facts could depend on the future. For the Ockhamist, it is not an open question whether facts like ‘JFK was shot’ or ‘The ants were in the yard’ depend on the future. But for the DEPENDENCE SOLUTION, this is an open question. So the (non- Ockhamist) DEPENDENCE SOLUTION rejects two features of Ockhamism: its method of determining dependence relations and, consequently, its limitation on which sort of facts could depend on the future.

---

References:


Kane, Robert. Free Will and Values (Albany: State University of New York Press, 1985)

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


