

UNIVERSITY OF CALIFORNIA

Santa Barbara

Defending Frankfurt: Moral Responsibility without Alternative Possibilities

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

by

Rick Lee Stoody

Committee in charge:

Professor C. Anthony Anderson, Co-Chair

Professor John Martin Fischer, University of California, Riverside, Co-Chair

Professor Aaron Zimmerman

September 2015

The dissertation of Rick Lee Stoody is approved.

---

Aaron Zimmerman

---

John Martin Fischer, Co-Chair

---

C. Anthony Anderson, Co-Chair

August 2015

To Jen, who helped make this, and so many other good things in my life, a reality.

## ACKNOWLEDGEMENTS

I thank the members of my committee (professors John Martin Fischer, C. Anthony Anderson, and Aaron Zimmerman) for their valuable feedback on many drafts of the chapters in this dissertation. I am grateful to the late Anthony Brueckner whose course “Freedom and Determinism” back in Spring 2009 kindled my interest in the present topic. I also wish to thank a few of the exceptional teachers I have had over the years: Dave Horner, J. P. Moreland, Doug Geivett, Garry Deweese, William Lane Craig, and Thomas Holden.

A word of thanks extends to my fellow graduate students in the UC Santa Barbara philosophy department, in particular: Timothy Linehan, Justin Clark, Dan Dolson, Christopher Cloos, Keith Hess, Joe Lee, and David Spewak. I thank them for their friendship and for many lively conversations.

Most importantly, I thank my family. My wife, Jennifer, to whom this dissertation is dedicated, has been a great encouragement during my seemingly unending graduate studies. She has had many of her own challenges to face during my studies, and she has borne them all with extraordinary grace. She is a treasure. My kids, Madeleine, Jack, and Emily, motivated me to get enough work done every day so we could play. Their love and tireless requests for a tussle have been just what I needed at the end of a long day. Each one of them has enriched my life beyond measure.

CURRICULUM VITA OF RICK LEE STOODY  
August 2015

AREAS OF SPECIALIZATION

Metaphysics, Philosophy of Action

AREAS OF COMPETENCE

Ethics, Epistemology, Logic

EDUCATION

B.S. Biochemistry, California State University, Fullerton, May 2005

M.A. Philosophy of Religion and Ethics, Biola University, May 2008

M.A. Philosophy, University of California, Santa Barbara, December 2011

Ph.D. Philosophy, University of California, Santa Barbara, August 2015

TEACHING EXPERIENCE

Teaching Associate: Introduction to Ethics, Critical Thinking (2012-2015)

Teaching Assistant: Ethics, Metaphysics, Late Modern Philosophy, Critical Thinking, Introduction to Philosophy, Introduction to Ethics (2009-2015)

PUBLICATIONS

“What’s It Like to Be a T. Rex?” In *Jurassic Park and Philosophy: The Truth is Terrifying*, edited by Nicolas Michaud and Jessica Watkins, 101-110. Chicago: Open Court, 2014.

PRESENTATIONS

“Undermining the Principle of Alternative Possibilities,” Pepperdine University, November 11, 2014.

“In Defense of Deterministic Frankfurt-Style Examples,” SoCal Philosophy Conference, University of California, San Diego, October 18, 2014.

“Defusing the Timing Objection to Frankfurt-Style Counterexamples,” SoCal Philosophy Conference, University of California, San Diego, October 20, 2013.

## AWARDS

2015: Graduate Student Association Excellence in Teaching Award 2015, university-wide award, UC Santa Barbara (Runner-up)

2013: 33rd annual Paul Wienpahl Award for Excellence in Teaching, UC Santa Barbara

2013: Interdisciplinary Humanities Center Graduate Collaborative Research Grant, UC Santa Barbara. Conference: Personhood, Possession and Place

2009-2015: University of California, Santa Barbara Doctoral Scholar TAship

2009-2015: R. W. Church Scholarship, University of California, Santa Barbara

2008-2009: University of California, Santa Barbara Doctoral Scholar Fellowship

2008: William W. Bass Memorial Scholarship Award, Biola University

## PROFESSIONAL MEMBERSHIPS

American Philosophical Association (2006 – Present)

Society of Christian Philosophers (2006 – Present)

## ABSTRACT

### Defending Frankfurt: Moral Responsibility without Alternative Possibilities

by

Rick Lee Stoody

According to the Principle of Alternative Possibilities (PAP), a person is morally responsible for what he has done only if he could have done otherwise. This principle, or something like it, seems quite plausible. And it has played an important role in arguments for the incompatibility of moral responsibility with causal determinism, divine foreknowledge, and divine providence. Yet, despite its intuitive appeal, PAP has not gone unchallenged. Beginning with Harry Frankfurt over forty years ago, a number of counterexamples (known as “Frankfurt examples”) have been offered. Since that time, a sizeable literature has developed around Frankfurt’s challenge.

The traditional strategy for attacking PAP is to attack it directly by attempting to provide a counterexample to the principle, as Frankfurt did. However, an alternative approach is to attack some other principle that PAP is connected to in an important way. I pursue the second strategy, targeting the following principle:

PAP(CC): When an agent is non-derivatively morally responsible for an action, he is so partly in virtue of having been able to have done otherwise.

I think the fundamental intuition behind an alternative possibilities requirement on moral responsibility is found in PAP(CC). Although PAP does not entail PAP(CC), if one rejects PAP(CC), then, absent other arguments in favor of PAP, one has little reason to accept PAP. I argue that PAP(CC) is false. Some Frankfurt examples—so-called “complete blockage” cases—are successful counterexamples to PAP(CC). However, since PAP does not entail PAP(CC), it remains possible that PAP is true even if PAP(CC) is false. For this reason, I consider two other arguments offered in support of PAP: the W-defense and the Deontic Maxim Defense. I argue that these arguments are unsuccessful. I conclude that even if PAP is in fact true, we have very little reason to believe that it is. After more than forty years of debate, it is time to say farewell to PAP.



## TABLE OF CONTENTS

Chapter 1. Alternative Possibilities and Moral Responsibility.....	1
1.1. Frankfurt's Counterexample.....	3
1.2. The Significance of Frankfurt's Argument .....	4
1.3. Refining the Principle of Alternative Possibilities .....	8
1.4. Summary of Chapters .....	12
Chapter 2. Undermining the Principle of Alternative Possibilities .....	15
2.1. A Prior Sign Frankfurt Example.....	15
2.2. The Flicker Defense.....	18
2.3. Grounding the Alternative Possibilities Requirement .....	22
2.4. Two Strategies For Targeting PAP(CC).....	25
2.5. The Dilemma Defense .....	27
2.6. The Deterministic Horn .....	29
2.7. Trouble for Fischer's Non-modal Strategy.....	31
2.8. Complete Blockage Frankfurt Examples.....	37
2.9. Concluding Remarks .....	43
Chapter 3. Indeterministic Frankfurt Examples.....	45
3.1. Occurrent Preemption Examples .....	46
3.2. Criticism of Occurrent Preemption Examples.....	49
3.3. Limited Blockage Examples.....	50
3.4. Criticism of Limited Blockage Examples .....	53
3.5. Internal Sign Examples.....	64
3.6. Criticism of Internal Sign Examples .....	67

3.7. Buffer Cases .....	69
3.8. Criticism of Buffer Cases .....	74
3.9. Concluding Remarks .....	83
Chapter 4. Two Arguments for the Principle of Alternative Possibilities .....	85
4.1. Widerker's Defense of PAP .....	85
4.2. Targeting the Irrelevance Principle .....	88
4.3. The W-defense.....	94
4.4. Replies to the W-defense .....	96
4.5. Justification for the Principle of Alternative Expectations.....	106
4.6. The Deontic Maxim Defense.....	113
4.7. Premise (D1): The Objective View of Moral Blameworthiness .....	114
4.8. Premise (D3): The Deontic Maxim .....	122
4.9. Concluding Remarks .....	129
References .....	130

## Chapter 1. Alternative Possibilities and Moral Responsibility

It is natural to think of ourselves as morally responsible for much of what we do. And being morally responsible for our actions seems to require that we have a particular sort of freedom or control over those actions.<sup>1</sup> But what sort of control? One natural response is to say that it is the *ability to do otherwise*—the power to either perform some action or to refrain from performing that action. Roderick Chisholm articulates this view,

Let us consider some deed, or misdeed, that may be attributed to a responsible agent: one man, say, shot another. If the man was responsible for what he did, then, I would urge, what was to happen at the time of the shooting was something that was entirely up to the man himself. There was a moment at which it was true, both that he could have fired the shot and also that he could have refrained from firing it. And if this is so, then, even though he did fire it, he could have done something else instead. (He didn't find himself firing the shot "against his will," as we say.) I think we can say, more generally, then, that if a man is responsible for a certain event or a certain state of affairs (in our example, the shooting of another man), then that event or state of affairs was brought about by some act of his, and the act was something that was in his power either to perform or not to perform. (Chisholm, 1964, p. 48)

Chisholm says that if a person is morally responsible for some action, then the action was entirely up to the person. And if the action was entirely up to the person, then he could have done something else—he could have refrained from performing that action. These two

---

<sup>1</sup> Some have challenged this claim, holding that what is required for moral responsibility is that the action be expressive of the attitudes and values of the agent. See, for instance, Thomas Scanlon (1998), Angela Smith (2005), and Nomy Arpaly (2006).

claims jointly entail what has come to be known as the “Principle of Alternative Possibilities”:<sup>2</sup>

PAP: If an agent is morally responsible for what she has done, then she could have done otherwise.

This principle, or something like it, seems quite plausible. After all, we generally don’t hold someone morally responsible for some action if we believe that she couldn’t have avoided doing what she did. If you are accused of some crime, say vehicular manslaughter, and you can convince your accuser that you could not have done otherwise—that it was not within your power *not* to drive over those pedestrians at the farmer’s market with your car (perhaps, through no fault of your own, your brakes went out and your steering wheel locked up)—then it seems as though you have done everything you need to do to absolve yourself of moral responsibility for the death of those pedestrians. It would be unreasonable, it seems, for your accuser to respond, “I recognize that you simply could not have done otherwise, but you are still blameworthy for driving your car over those people.”<sup>3</sup>

Despite its intuitive appeal, PAP has not gone unchallenged. Beginning with Harry Frankfurt (1969) over forty years ago, a number of counterexamples (known as “Frankfurt examples”) have been offered, and a sizable literature has developed around those counterexamples. After all this time there is no clear consensus on whether any of these counterexamples are successful. In this dissertation I shall argue that Frankfurt examples

---

<sup>2</sup> Harry Frankfurt (1969) named it the “principle of *alternate* possibilities.”

<sup>3</sup> Of course, things are a bit more complicated here. If the reason you are unable to refrain from evading your taxes at a particular moment is that you arranged things earlier so that you wouldn’t be able to do otherwise at this moment, then it seems that you are still morally responsible for evading your taxes even though you couldn’t do otherwise at this particular moment. I will have more to say about these sorts of cases later when I clarify what is meant by PAP.

provide us with good reason to doubt the truth of PAP, and moreover, that arguments provided by PAP defenders in support of PAP are unsuccessful. Along the way I shall explain why consensus on the success of Frankfurt examples has been elusive thus far.

In this chapter I shall lay the groundwork for what follows in this dissertation. I begin by sketching Frankfurt's argument against PAP and explaining its significance to debates involving moral responsibility. Then I clarify how we should understand the principle, offering a fully specified version of PAP. I conclude with a summary of the remaining chapters.

### ***1.1. Frankfurt's Counterexample***

I shall have more to say about Frankfurt's argument in Chapter 2. For now, a brief sketch shall suffice. In his influential article, "Alternate Possibilities and Moral Responsibility," Frankfurt says that the typical examples offered in support of PAP—cases of coercion or manipulation, for instance—are situations in which the circumstances that prevent an agent from doing otherwise are the *very same* circumstances that bring it about that the agent performs the action he does. But Frankfurt offers a situation in which the circumstances which ensure that an agent is unable to do otherwise *do not themselves* bring about his performing the action. Here is Frankfurt's example:

Suppose someone—Black, let us say—wants Jones to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until Jones is about to make up his mind what to do, and he does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones is going to decide to do something *other* than what he wants him to do. If it does become clear that Jones is going to decide to do something else, Black takes effective steps to ensure that Jones decides to do, and that he does do, what he wants him

to do. Whatever Jones's initial preferences and inclinations, then, Black will have his way...Now suppose that Black never has to show his hand because Jones, for reasons of his own, decides to perform and does perform the very action Black wants him to perform. In that case, it seems clear, Jones will bear precisely the same moral responsibility for what he does as he would have borne if Black had not been ready to take steps to ensure that he do it. (Frankfurt, 1969, pp. 835-836)

Let us suppose that Jones performs some action for which he would be morally responsible under normal circumstances (i.e., those in which he could have done otherwise). Frankfurt says that since Jones performs the action that he does for reasons of his own, and not *because* of his circumstances, Jones is morally responsible for what he has done, despite the fact that his circumstances precluded him from doing otherwise. Therefore, PAP is false.

### ***1.2. The Significance of Frankfurt's Argument***

Frankfurt's argument against PAP has some important implications. Before discussing them, it will be helpful to briefly define some key terms. *Causal determinism* is the thesis that given the past and the laws of nature, there is only one physically possible future.<sup>4</sup> One question that has received much attention is whether causal determinism is compatible with

---

<sup>4</sup> See van Inwagen (1983, p. 65). This is the way causal determinism is typically defined, and I will go along with it. However, I think this way of defining causal determinism is a bit misleading. Derk Pereboom makes the following distinction:

[T]here are two ways of thinking about causal determinism: (a) as involving the claim that events are *entailed* by propositions that describe preceding conditions and the laws of nature, and (b) as involving the claim that events are actually *produced* by such preceding conditions in accord with the laws of nature. (Pereboom, 2012, p. 305)

I think (b) is preferable to (a). It seems to capture what we mean by *causal* determinism better than (a). But nothing I conclude in this dissertation will hinge on whether one adopts (a) or (b).

moral responsibility.<sup>5</sup> *Compatibilists*, as I shall use the term, are those who answer “yes” to this question. *Incompatibilists* are those who answer “no.”

A classical argument for incompatibilism goes like this:

- (1) If an agent is morally responsible for what she has done, then she could have done otherwise (PAP).
- (2) If causal determinism is true, then no agent could have done otherwise.
- (3) Therefore, if causal determinism is true, then no agent is morally responsible for what she has done.

Before Frankfurt’s article, both compatibilists and incompatibilists generally agreed that premise (1) was true. What was open for debate was premise (2). The typical compatibilist response to premise (2) was to analyze the phrase ‘could have done otherwise’ as a counterfactual conditional. On this “conditional analysis,” ‘could have done otherwise’ is understood to mean that *if an agent had wanted to do otherwise, he would have done otherwise*. Since causal determinism is a thesis about what will happen in the future, given the (actual) past, causal determinism is consistent with there being a different future, given a different past. It is now generally agreed that this compatibilist defense of (2) is mistaken.<sup>6</sup>

---

<sup>5</sup> A related question is whether free will is compatible with causal determinism. If free will is taken to be a necessary condition for moral responsibility, then one’s answer to each question will be the same. Some philosophers have held that moral responsibility is compatible with causal determinism while allowing that free will may be ruled out by causal determinism (e.g., see Fischer, 1994; 1998; Frankfurt, 1971). Note that this position, which John Martin Fischer calls “semi-compatibilism,” is a form of compatibilism given the way I will be using the term.

<sup>6</sup> For some classic objections see J. L. Austin (1956), Chisholm (1964), and Keith Lehrer (1968).

A powerful argument in support of (2), first developed by Carl Ginet (1966), came to be known as the “Consequence Argument.” Here is a brief and informal sketch of the argument, offered by Peter van Inwagen:

If [causal] determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born; and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our own acts) are not up to us. (van Inwagen, 1983, p. 16)

The goal of the argument is to demonstrate that if causal determinism is true, then for any action performed by any agent, no agent has or ever had the ability to do otherwise. In saying that the laws of nature and events in the remote past “are not up to us,” van Inwagen is saying that there is nothing we can do now to alter these things; they are outside of our control. And if this is true, then given causal determinism, we are unable to do otherwise.<sup>7</sup>

The strong appeal of the Consequence Argument, along with the assumption of PAP, and difficulties with the conditional analysis, left traditional compatibilists with the difficult task of defending a position that seems implausible. However, with the arrival of Frankfurt’s challenge to premise (1), compatibilists could grant that causal determinism

---

<sup>7</sup> Those incompatibilists who also hold that some people are morally responsible for what they do will need to go on to show how moral responsibility is compatible with causal *indeterminism*. There is a famous line of reasoning (which mirrors the Consequence Argument) for the conclusion that causal indeterminism rules out the sort of control required for moral responsibility:

If causal determinism is false, then our acts are the indeterministic consequences of the laws of nature and the remote past. But it is not up to us what indeterministic consequences result from the laws of nature; neither it is up to us what indeterministic consequences result from events of the remote past. Therefore, the consequences of these things (including our own acts) are not up to us.



rules out the ability to do otherwise while maintaining that it does not rule out moral responsibility. For this reason, many compatibilists now prefer to challenge PAP.<sup>8</sup>

It is important to keep in mind that Frankfurt's argument against PAP is not an argument *for* compatibilism.<sup>9</sup> If successful, it enables compatibilists defeat the classical argument for incompatibilism without having to take the Consequence Argument head-on. However, the classical argument is just one argument for the incompatibility of moral responsibility with causal determinism. Other arguments have been offered that do not depend on PAP.<sup>10</sup>

Frankfurt's argument against PAP also has implications for longstanding theological debates over whether infallible divine foreknowledge is compatible with human responsibility, and whether divine providence is compatible with human responsibility. Some have thought that divine foreknowledge threatens the possibility of human responsibility because if God knows the entire future, and cannot be mistaken, then nothing—including the way each human being acts—could happen differently than it does. If this is true, and if human responsibility requires the ability to do otherwise, then no human

---

<sup>8</sup> Several compatibilists have not gone this route. See, for instance, Bernard Berofsky (2002), Kadri Vihvelin (2004), Michael Fara (2008), Michael Smith (2003), Daniel Cohen and Toby Handfield (2007).

<sup>9</sup> Fischer (1982, p. 34) was the first to point out that incompatibilists might reject PAP while maintaining that causal determinism eliminates moral responsibility.

<sup>10</sup> PAP concerns what may be called a "leeway condition" on moral responsibility. However, some philosophers have defended a "sourcehood condition" on moral responsibility: an agent is morally responsible for his action only if he is the *ultimate source* or *originator* of that action. What does it mean to be the ultimate source of one's actions? Very roughly, it is to be "self-determining"; it is for our actions to be "up to us" and not the result of external forces (O'Connor, 1996, p. 143). An argument for incompatibilism can be developed around the sourcehood condition by arguing that causal determinism rules out ultimate sourcehood or origination. See, for example, Laura Ekstrom (1998), David Hunt (2005), Robert Kane (1996), Derk Pereboom (2001, 2014), Eleonore Stump (1996), Kevin Timpe (2008), and Linda Zagzebski (2000).

being is morally responsible for anything she does. Similarly, divine providence has been thought to threaten human responsibility because if God determines everything that happens, then human beings lack the ability to behave differently than they do. However, if Frankfurt is correct and PAP is false, then neither of these arguments goes through.

### ***1.3. Refining the Principle of Alternative Possibilities***

So far, I have been relying on an intuitive understanding of PAP. In this section I want to clarify exactly what I mean by the principle. I shall begin with moral responsibility. There are many things that people may be morally responsible for: their actions, omissions, beliefs, attitudes, desires, etc. I shall limit PAP to the claim that moral responsibility for an action requires alternative possible courses of action at the time at which the action for which a given agent is putatively responsible is initiated. If decisions (or the adopting of plans or the execution of judgments) are themselves mental actions, responsibility for such is also falls within the scope of the principle (so interpreted).

What is it for an agent to be morally responsible for some action? I like the way Derk Pereboom puts it:

[F]or an agent to be *morally responsible for an action* is for this action to belong to the agent in such a way that she would deserve blame if the action were morally wrong, and she would deserve credit or perhaps praise if it were morally exemplary...This characterization leaves room for an agent's being morally responsible for an action even if she does not deserve blame, credit, or praise for it—if for example, the action is morally indifferent. (Pereboom, 2001, p. xx)

Moral responsibility is the sort of responsibility that is required for praise or blame that is deserved.<sup>11</sup> It is distinct from mere causal responsibility. A boulder that destroys someone's house in a mudslide is causally responsible for the destruction of the house even though it is not morally responsible. While both agents and non-agents alike can be causally responsible for events, only agents can be morally responsible for them. Additionally, agents can be causally responsible for events without being morally responsible for them. While rushing to rescue a drowning child I might unintentionally bump someone, causing her drink to spill. In such a situation, I may be causally responsible for the drink spilling even if I am not morally responsible.

In what follows, I shall focus on cases of moral blameworthiness.<sup>12</sup> This is, in part, for two reasons. First, I agree with what Gary Watson has to say about focusing on blame:

We seem to have a richer vocabulary of blame than of praise. This slant is not due solely to mean-spiritedness. At least part of the explanation is that blaming tends to be a much more serious affair; reputation, liberty, and even life can be at stake, and understandably we are more concerned with the conditions of adverse treatment than with those of favorable treatment. (Watson, 2004, p. 283)

Because of how serious moral blame can be, the most powerful intuitive responses are elicited by considering cases involving blameworthiness. Second, some philosophers think that there is an important asymmetry between moral praiseworthiness and blameworthiness. They hold that one requirement on an agent's being morally blameworthy for performing an

---

<sup>11</sup> Several influential accounts of moral responsibility have been offered (see, for example, Oshana, 1997; Strawson, 1962; M. J. Zimmerman, 1988, p. 38). Nothing I shall say in this dissertation hinges on the truth of any particular account.

<sup>12</sup> Several theories of what it is for an agent to be morally blameworthy have been offered. For an excellent overview of these theories see Justin Coates and Neal Tognazzini (2013). I will not presuppose any particular theory.

action is that he could have done otherwise, but that there is no such requirement for an agent to be morally praiseworthy.<sup>13</sup> I wish to avoid this debate, so I shall restrict moral responsibility in PAP to the sort of responsibility that is required for moral blameworthiness.

I shall further limit moral responsibility to *non-derivative* (or *direct*) moral responsibility. A person is non-derivatively (or directly) morally responsible for performing an action if he is morally responsible for the action in virtue of being morally responsible for some other state of affairs (e.g., for performing some other action at an earlier time). Were PAP not limited in this way, counterexamples would be quite easy to come by. Consider, for instance, a drunk driver who is so drunk that by the time he sees a car braking in front of him, no matter what he does, he could not avoid hitting that car. He seems to be morally responsible for hitting the car, yet he could not have done otherwise. But if we limit PAP to non-derivative responsibility, cases like this will not be counterexamples. The drunk driver's moral responsibility for hitting the car is derivative of his moral responsibility for his earlier decision to drive while drunk (on the assumption that he needn't have made this decision nor performed the action he then decided to perform).

Finally, I turn to how to interpret 'could have done otherwise.' One natural way of reading this phrase is 'could have performed a different action.' On this interpretation of PAP, an agent is morally responsible for doing something only if he could have performed some other action instead. This would preclude cases of moral responsibility where an agent's only alternative is to simply refrain from (or omit) performing any action.<sup>14</sup> However, I see no principled reason for restricting PAP in this way. I shall take 'could have

---

<sup>13</sup> See, for instance, Dana Nelkin (2008) and Susan Wolf (1990, pp. 79-81).

<sup>14</sup> See Randolph Clarke (2014) for examples.

done otherwise' to mean 'could have performed some other action or could have refrained from performing the action,' where refraining from performing an action is not itself an action. Furthermore, I shall read 'could have done otherwise' synchronically: the time at which the agent performs the action for which he is morally responsible is the same time at which the agent could have done otherwise.<sup>15, 16</sup>

How should we interpret 'could'? Claims about what an agent *could* have done might be taken to be claims about what it is *possible* for the agent to have done. Alternatively, they might be understood to be claims about what an agent is *able* to have done (i.e., claims about what an agent had it within his power to do). I shall treat PAP as a claim about ability.

Given the preceding, the resulting fully specified formulation of PAP is as follows:

---

<sup>15</sup> See Alfred Mele (2006, pp. 84-86) for a discussion of synchronic and diachronic readings of PAP. Rather than limiting PAP to non-derivative responsibility and interpreting it synchronically, PAP can be interpreted diachronically: "A person is morally responsible for what he did at *t* only if he was able...either before *t* or at *t*, to do otherwise at *t*" (Mele, 2006, p. 85). On the diachronic interpretation, there is no need to limit PAP to non-derivative moral responsibility. I do not think anything of consequence for this dissertation turns on which interpretation one chooses. So, for simplicity's sake, I shall discuss PAP in terms of the synchronic reading. If the reader prefers the diachronic reading, she is welcome to translate what I say to fit that reading.

<sup>16</sup> A quick word about the timing of actions is in order here. Some actions (perhaps all actions) are extended over time. The act of fixing a clogged sink is not momentary (alas), but extends through time. Once an agent has *initiated* the act of unclogging the sink, he may refrain from unclogging the sink by stopping at any moment before the sink is unclogged. For those actions that are not instantaneous, a natural reading of 'the time at which the agent performs the action...is the same time at which the agent could have done otherwise' is 'the time at which the agent *begins* or *initiates* the action is the same as the time at which the agent could have done otherwise than *begin* or *initiate* the action.' I shall treat the actions (and decisions) discussed in this dissertation as actions that are performed at an instant. However, readers that prefer to treat these actions (and decisions) as actions that are extended through time may replace 'the time at which the agent performs the action' with 'the time at which the agent initiates the action.'

Necessarily, if an agent S is non-derivatively morally responsible for performing some action A at time t, then S was able to (i) perform some action other than A at t or (ii) refrain from performing A at t.

Throughout this dissertation, whenever I refer to PAP, this is the principle I shall be referring to.

#### ***1.4. Summary of Chapters***

This dissertation is focused on one central question regarding moral responsibility: must an agent be able to do otherwise for her to be morally responsible for what she has done? I argue that we have very little reason to answer “yes” to this question.

In Chapter 2, I argue that the primary motivation for PAP is found in another principle—PAP(CC)—and that this principle is false. One problem with targeting PAP directly is that Frankfurt examples face a dilemma. If the circumstances of the scenario eliminate every one of the agent’s alternatives, then the example is susceptible to the charge that the circumstances amount to causal determinism and that it would be question begging to claim that an agent in such circumstances is morally blameworthy. However, if the circumstances do not eliminate every alternative, then the agent could have done otherwise. Either way, it is argued, the Frankfurt example fails.

Rather than target PAP directly, I target another principle:

PAP(CC): When an agent is non-derivatively morally responsible for an action, he is so partly in virtue of having been able to have done otherwise.

There is little reason to accept PAP if one rejects PAP(CC). And I argue that there is good reason to reject PAP(CC). There are two strategies for using Frankfurt examples against PAP(CC). One is to offer them as metaphysically possible scenarios in which an agent is morally responsible and has access to *some* alternative possibilities yet lacks access to

*robust* alternatives—i.e., alternative possibilities that could be explanatorily relevant to an agent’s moral responsibility. Call this the “modal strategy.” The other strategy is to use Frankfurt examples as a way of showing that robust alternatives are not part of the conceptual fabric of moral responsibility. Call this the “non-modal” strategy. In the remainder of Chapter 2, I pursue the non-modal strategy against PAP(CC). In particular, I argue that what are called “complete blockage” Frankfurt examples can be used in a successful argument against PAP(CC) even if the elimination of alternative possibilities in these examples amounts to casual determinism. I conclude that PAP(CC) is false.

In Chapter 3, I consider four leading approaches to designing indeterministic Frankfurt examples to be used in the modal strategy against PAP(CC). I argue that, unfortunately, none of them are successful. I think this partly explains why it remains controversial whether Frankfurt examples are successful, even after more than forty years of debate. The failure to provide a clear-cut modal counterexample to PAP(CC) has led some to conclude that Frankfurt’s argument fails. However, as I argue in Chapter 2, Frankfurt examples do not need to succeed in a modal strategy against PAP(CC) in order to succeed in demonstrating that PAP(CC) is false.

Since PAP does not entail PAP(CC), it remains possible that PAP is true even if PAP(CC) is false. In Chapter 4, I consider and reject two arguments in support of PAP. The first argument hinges on the intuitively plausible claim that an agent is morally blameworthy for performing some action only if it would have been reasonable to expect the agent not to have performed that action (call this principle “PAE”). If we also assume that it would be reasonable to expect an agent not to do something only if he could have avoided doing it, then PAP logically follows. I argue that there is good reason to reject PAE in favor of a competing principle that does not support PAP. That principle is the “Quality of Will

Thesis”: when an agent is morally responsible for performing some action, he is so in virtue of the quality of the agent’s will that is revealed by the performance of that action.

The second argument for PAP involves Kant’s maxim that “ought” implies “can.” If we also grant that an agent is morally blameworthy for performing some action only if he ought not to have performed that action, then Kant’s maxim entails PAP. I argue that there is good reason to reject the maxim. I conclude that in the absence of other arguments in favor of PAP, we should be skeptical of its truth. For all we know, PAP *could* be true, but we have little reason to believe that it *is* true.



## **Chapter 2. Undermining the Principle of Alternative Possibilities**

Frankfurt examples have traditionally been offered as direct counterexamples to the Principle of Alternative Possibilities:

PAP: Necessarily, an agent is non-derivatively morally responsible for what she has done only if she could have done otherwise.

On this approach, a metaphysically possible situation is described in which an agent is non-derivatively morally responsible for performing some action even though she could not have done otherwise. After considering this situation we are supposed to conclude that PAP is false. One difficulty with this approach, however, is that some have claimed that the elimination of every alternative possibility amounts to causal determinism. So, to claim that an agent who lacks alternative possibilities is morally responsible begs the question against incompatibilists.

In this chapter I shall pursue a different strategy. Rather than target PAP directly, I shall target another principle:

PAP(CC): When an agent is non-derivatively morally responsible for an action, he is so partly in virtue of having been able to have done otherwise.

There are two strategies for targeting PAP(CC), a modal strategy and a non-modal strategy. In this chapter I shall focus on the one I think is successful, the non-modal strategy. In the next chapter I shall focus on the modal strategy. I conclude that PAP(CC) is false, and therefore, absent other arguments in support of PAP, there is little reason to accept it.

### ***2.1. A Prior Sign Frankfurt Example***

In Harry Frankfurt's (1969) influential article, "Alternative Possibilities and Moral Responsibility," he attempts to provide a counterexample to PAP by constructing a

metaphysically possible situation in which an agent is non-derivatively morally responsible for what he did *even though* his circumstances prevented him from refraining from acting as he did.<sup>1</sup> Frankfurt believed that these circumstances could be designed in such a way that, although they prevented an agent from accessing alternatives to performing the action, they did not in any way bring about his performing the action. To see how this might go, consider the following Frankfurt example, Bribe:

Smith is pushing a bill through Congress that puts severe restrictions on neuroscientific research. Congressman Jones has been offered a large sum of money by the League of Nefarious Neuroscientists (LNN) to vote against the bill. Black, the founder of the LNN, would prefer that Jones accept the bribe on his own. However, not wanting to take any chances, Black has, unbeknownst to Jones, implanted a special device into Jones's brain. Lucky for Black, Jones has a tell that reliably indicates what he will do. If Jones is going to decide to accept the bribe on his own by a certain time,  $t_3$ , he will exhibit an involuntary facial twitch,  $F$ , prior to  $t_1$ . If Jones exhibits  $F$  prior to  $t_1$ , Black will not activate the device. However, if Jones is not going to decide to accept the bribe on his own by  $t_3$ , he will not display  $F$  prior to  $t_1$ , and Black will activate the device at  $t_2$ , causing Jones to choose to accept the bribe by  $t_3$ . As things actually happen, Jones exhibits  $F$  before  $t_1$ , Black never activates the neural device, and Jones decides to accept the bribe by  $t_3$ .<sup>2</sup>

---

<sup>1</sup> In what follows, unless otherwise noted, I shall use "morally responsible" to mean *non-derivatively* morally responsible. Recall from Chapter 1 that by non-derivatively morally responsible I mean, very roughly, that a person is morally responsible for an event or state of affairs and his being responsible does not depend on his being morally responsible for some other event or state of affairs.

<sup>2</sup> Frankfurt suggests introducing a prior sign to the examples in a footnote of his original (1969) article. John Martin Fischer (1999, 2003) provides Frankfurt examples where the prior sign is made explicit.

A key feature of Frankfurt examples like this one is that they involve a prior sign that indicates what the agent is going to do (unless some outside force intervenes). For this reason, these sorts of cases are called “prior sign” Frankfurt examples.

It is important to notice that although the circumstances prevent Jones from failing to decide to accept the bribe, they do not actually bring it about that Jones decides to accept the bribe. As Frankfurt says, Black and his device “played no role at all in leading him to act as he did” (1969, p. 836). Jones did not decide to accept the bribe “*because* he could not have done otherwise” (Frankfurt, 1969, p. 837). Frankfurt goes on,

When a fact is in this way irrelevant to the problem of accounting for a person’s action it seems quite gratuitous to assign it any weight in the assessment of his moral responsibility. Why should the fact be considered in reaching a moral judgment concerning the person when it does not help in any way to understand...what made him act as he did? (Frankfurt, 1969, p. 837)

Thus, it seems that Jones is morally responsible for deciding to accept the bribe by t3 (provided the other requirements for moral responsibility are met).<sup>3</sup> Yet, Jones could not have done otherwise. So Bribe and other prior sign Frankfurt examples like it appear to show that PAP is false.

---

<sup>3</sup> In Frankfurt’s original article he provides two lines of argument for this conclusion, though he does not clearly differentiate them. The argument quoted above is one of them. For a discussion of these two distinct arguments, see Widerker (2000, pp. 188-191).

## 2.2. *The Flicker Defense*

Some critics of Frankfurt's argument have attempted to defend PAP from cases like Bribe by employing what has come to be called the "flicker defense."<sup>4</sup> The proponent of the flicker defense (the "flicker defender") grants that the agent does not have the usual alternatives available, but she attempts to find some other alternative possibility available to the agent that has been overlooked. One way of doing this, in response to Bribe, is to grant that Jones could not do otherwise than decide to accept the bribe but to say that this is not what Jones is morally responsible for. Rather, what he is morally responsible for is for deciding to accept the bribe *on his own* without interference from Black. And if this is what Jones is morally responsible for, then it turns out that he did have an alternative available. He could have brought it about that he was coerced by black to decide to accept the bribe, instead of deciding to accept it on his own. And, had he been coerced, then he would not have been morally responsible for his decision.

The problem with this response, however, is that it puts the flicker defender in the unfortunate position of claiming that if Black does not intervene, Jones is morally responsible for deciding to accept the bribe on his own but is *not* morally responsible for deciding to accept the bribe. But this is implausible. If an agent is morally responsible for A-ing on his own, then he is morally responsible for A-ing. As Robert Kane, a prominent proponent of PAP who rejects this version of the flicker defense, puts it, "That was the point the Frankfurt-type examples were trying to make in the first place: we are responsible for the things we do on our own when no one interferes" (Kane, 2005, p. 86). Thus, if Jones is morally responsible for deciding to accept the bribe on his own, then he is morally

---

<sup>4</sup> The name comes from John Martin Fischer (2003). He considers and responds to four different versions of the flicker defense. Here I consider only the first and the fourth versions.

responsible for deciding to accept the bribe. But he could not do otherwise than decide to accept the bribe. So Frankfurt's argument against PAP seems to hold up against this version of the flicker defense.

Another approach the flicker defender might take comes from tracing the agent's action back to the prior sign. As things stand in Bribe, Jones exhibits the prior sign, F, which indicates to Black that he does not need to intervene. Given the presence of Black and his device, Jones could not do otherwise than decide to accept the bribe. However, when we trace Jones's action back to the prior sign, it looks as though he did have an alternative possibility available after all: he might not have exhibited F. Had Jones not exhibited F, he would not have been morally responsible for his decision since Black would have activated the device and he would not have made the decision on his own. Thus, the flicker defender might attempt to trace Jones's moral responsibility for his decision back to this alternative possibility, this "flicker of freedom" as John Martin Fischer (2003) calls it. Although Jones could not do otherwise than decide to accept the bribe, he could have failed to exhibit F. Thus, the flicker defender might claim that Jones's moral responsibility for his decision is derivative of his moral responsibility for exhibiting F, something he had an alternative to.

It does not seem possible to eliminate *every* alternative possibility in Frankfurt examples like Bribe since they are built around the presence of a prior sign that indicates what the agent is going to choose to do on his own. Thus, it looks like there will always be *some* alternative that the flicker defender could claim is what grounds the agent's moral responsibility. However, Fischer argues that it is unreasonable to think that Jones's moral responsibility for his decision is derivative of his being morally responsible for exhibiting the prior sign. The problem is that Jones's exhibiting F—a facial twitch—is *involuntary*,

something over which he does not have control.<sup>5</sup> It is implausible that the involuntary display of the prior sign could make the difference between whether Jones is morally responsible for his decision or not. Fischer makes the point this way,

[I]magine, just for a moment, that there are absolutely no alternative possibilities, even the flimsy and exiguous flickers of freedom we have recently been entertaining. An alternative-possibilities control theorist would say that under such circumstances the relevant agent cannot be morally responsible for his choice and action. Now add the flickers of freedom we have been considering...I find it very hard to see how adding this power can transform a situation in which there is no moral responsibility into one in which there is moral responsibility. (Fischer, 2002, p. 289)

Since Jones's exhibiting the prior sign is involuntary, it does not seem to be something he could be morally responsible for. Fischer says,

I contend that the mere involuntary display of some sign—such as a neurological pattern in the brain, a blush, or a furrowed brow—is too thin a reed on which to rest moral responsibility. The power involuntarily to exhibit a different sign seems to me to be insufficiently robust to ground our attributions of moral responsibility. (Fischer, 2002, p. 289)

Intuitively, the alternatives available must be relevant to the agent's moral responsibility.

That is, they must be, as Fischer says, "robust;" they must be alternatives in virtue of which an agent could be held morally responsible. A natural condition on such alternatives is that

---

<sup>5</sup> The flicker defender might argue that the prior sign must be *voluntary*. However, this won't help. Suppose that, instead of an *involuntary* action like a facial twitch, the prior sign is one Jones could *voluntarily* exhibit, say raising his right arm. The Frankfurt defender could alter Bribe so that Black focuses on controlling this earlier voluntary action and Frankfurt's argument could be shifted back to this earlier act. If Jones is going to voluntarily raise his right arm which leads to his deciding to accept the Bribe, then Black will not interfere. Otherwise, Black intervenes, ensuring that Jones raises his right arm.

they be under the agent's *voluntary* control.<sup>6</sup> After all, there are many ways things could have gone such that Jones did not decide to accept the bribe. He could have been struck dead, he could have suddenly developed a case of amnesia, he could have been arrested, etc. All of these are metaphysical possibilities. But surely it is not the case that the mere availability of these sorts of possibilities is what makes the difference between whether Jones is morally responsible for his decision or not. Likewise, it cannot be the case that Jones's moral responsibility for deciding to accept the bribe hinges on the possibility that he does not perform an involuntary act. On this point, Fischer remarks,

The existence of *various* genuinely open pathways is alleged [by PAP defenders] to be *crucial* to the idea that one has *control* of a relevant kind. But if this is so, I suggest that it would be very puzzling and unnatural to suppose that it is the existence of various alternative pathways along which one does *not* act freely that shows that one has control of the kind in question. How exactly *could* the existence of various alternative pathways along which the agent does *not* act freely render it true that the agent has the relevant kind of control? (Fischer, 2003, p. 35)

Fischer argues that the PAP defender must do more than simply identify alternative possibilities in the Frankfurt examples. He must also show that these alternatives are relevant to the agent's moral responsibility, that they are the sorts of alternatives that could ground an agent's moral responsibility. But these mere flickers of freedom don't seem to be

---

<sup>6</sup> Typically, in addition to this volitional condition on robust alternatives, there is an epistemic condition: an agent must have some sort of understanding of the moral significance of the alternative available. For example, suppose that, unbeknownst to Jones, his coffee was laced with a new chemical that prevents people from accepting bribes. As it happens, Jones never sips his coffee and decides to accept the bribe. Although Jones could have avoided accepting the bribe by sipping his coffee, that is not a robust alternative since he is unaware of this fact. For a recent discussion of the requirements for robust alternatives see Pereboom (2009).

able to meet that requirement. It would be quite strange to say that Jones is morally responsible for deciding to accept the bribe *because* he might not have involuntarily twitched. Such alternatives are, as Fischer says, too “flimsy and exiguous” to play a part in grounding an agent’s moral responsibility. Thus, although these sorts of alternatives cannot be entirely eliminated from prior sign Frankfurt examples, it is implausible to think that it is in virtue of having access to these alternatives that the principal agent is morally responsible for performing an action that he could not have avoided performing.

### ***2.3. Grounding the Alternative Possibilities Requirement***

As presented, PAP does not explicitly require a deep conceptual connection between moral responsibility and alternative possibilities. PAP says that in every possible world in which someone is morally responsible for performing some action at a particular time, that person also has an alternative to performing that action at that time. However, an alternative possibilities requirement on moral responsibility would be unmotivated or unprincipled if it did not require alternative possibilities that could explain (in part) why a given agent is morally responsible for his action. Flickers of freedom are not the sort of alternatives that could perform that function. The intuitive plausibility of PAP seems to rest on there being a deep conceptual connection between alternative possibilities and moral responsibility. After all, if alternative possibilities are explanatorily irrelevant to moral responsibility, why think they are necessary?<sup>7</sup> The most natural reason for accepting PAP is that the availability of

---

<sup>7</sup> Of course, it is open for the PAP defender to dig in his heels and claim that PAP needs no support. On such a view, one might contend that it is reasonable to hold to the truth of PAP unless it is shown to be false. What I have to say will not sway those who take this position. It is not incoherent in any clear sense. However, one of the things that Frankfurt’s argument shows us is that an alternative-possibilities framework for moral responsibility has a clear competitor in an actual-sequence framework. (I shall say more about this in Chapter 4.) Thus, one might wonder which framework one ought to adopt. Although the PAP defender could simply assert that the alternative-possibilities framework is the one that



alternative possibilities is an integral part of what it is to be morally responsible. And the only alternatives that could play that role are robust alternatives. As Fischer says,

The intuitive picture behind the alternative-possibilities control requirement is that moral responsibility requires that the agent select one from among various genuinely open paths the world might take. There are two important ideas here. One is that there must be various paths genuinely available to the agent (at least at some times suitably related to the time of the behavior under consideration). The second idea is that the agent (and not some outside force or mere chance) selects which path will be the path into the future. It seems to me that both ideas are important components of the traditional conception of the sort of control associated with moral responsibility—alternative-possibilities control. (Fischer, 1999, p. 99)

The fundamental intuition that lies behind an alternative possibilities requirement on moral responsibility is that moral responsibility requires a certain sort of control, and this sort of control is not present when the only alternatives an agent has access to are non-robust.

Frankfurt makes a similar point:

The critical issue concerning PAP, then, is not whether it is always possible that an agent who is morally responsible for performing a certain action might have acted differently. Rather, it is whether that possibility—even assuming that it is real—counts for anything in determining whether he is morally responsible for what he did. In my view, neither the avoidability of an action nor its unavoidability is morally significant when it has nothing to do with how the action came to be performed. (Frankfurt, 2003, p. 340)

---

ought to be adopted, such a refusal to explain why seems to be unprincipled, and will not persuade those who genuinely wonder whether it ought to be adopted over the alternative.

This way of motivating an alternative possibilities requirement on moral responsibility is expressed by the following principle:

PAP(CC): When an agent is non-derivatively morally responsible for an action, he is so partly in virtue of having been able to have done otherwise.<sup>8</sup>

PAP(CC) states that having alternative possibilities is part of the “conceptual fabric” or “metaphysical nature” of moral responsibility. What it is to be morally responsible for performing some action is (in part) to have access to alternatives to performing that action. The fundamental intuition behind an alternative possibilities requirement on moral responsibility is found in PAP(CC), not PAP.<sup>9</sup>

To see the distinction between PAP and PAP(CC) more clearly, consider the following example.<sup>10</sup> Let  $P$  be some proposition.  $(P \vee \sim P)$  is a necessary condition for an agent’s being morally responsible for some action. That is,  $(P \vee \sim P)$  will be true in every possible world where an agent is morally responsible.<sup>11</sup> But this tautology has no role to play in explaining *why* the agent is morally responsible for his action. It would be a mistake to think that an agent is morally responsible partly in virtue of the fact that  $(P \vee \sim P)$  is true. That tautology does not enter into what it is to be morally responsible. Likewise, for PAP to

---

<sup>8</sup> “CC” is for “constituent conditional.” Drawing from Felipe Leon and Neal Tognazzini (2010), this formulation of PAP is intended to state that the ability to do otherwise is a constituent property of non-derivative moral responsibility, and not merely a property that happens to show up in every possible world whenever non-derivative responsibility does.

<sup>9</sup> Michael McKenna argues that PAP is “too inclusive” since it “can be satisfied even if the only options available to an agent are morally (that is, deliberately) insignificant ones” (McKenna, 2003, pp. 208-209).

<sup>10</sup> This example is inspired by one offered by Christopher Franklin (2011a).

<sup>11</sup> Of course,  $(P \vee \sim P)$  will also be true in possible worlds where agents are not morally responsible. However, so will the proposition that an agent could have done otherwise. It is agreed on all sides of the debate over PAP that the ability to do otherwise is not sufficient for moral responsibility.

be true, the ability to do otherwise need not have any role to play in accounting for *why* an agent is morally responsible. But this is not so for PAP(CC). PAP(CC) states that there is a deep conceptual or metaphysical connection between the ability to do otherwise and moral responsibility.

If we can establish the falsity of PAP(CC), then, absent other arguments in favor of PAP, one has little reason to accept PAP. Although it is possible that PAP is true even if PAP(CC) is false, it is difficult to see why we should think it is.<sup>12</sup> In the remainder of this chapter I shall focus on arguments against PAP(CC).

#### ***2.4. Two Strategies For Targeting PAP(CC)***

Shifting the discussion from PAP to PAP(CC) opens up new avenues for employing Frankfurt examples. The traditional approach, which directly targeted PAP, was to provide a modal counterexample, demonstrating that it is metaphysically possible for an agent to be morally responsible for performing some action while lacking the ability to do otherwise. Since PAP is a claim about metaphysical necessity, this is the only way to show that PAP is false. This general strategy can also be applied to PAP(CC). If one can provide a metaphysically possible scenario in which an agent is morally responsible for some action yet lacks access to any alternative possibilities that might be explanatorily relevant to the agent's moral responsibility, then PAP(CC) is false. Call this the "modal strategy." Since non-robust alternatives are not explanatorily relevant to an agent's moral responsibility, a Frankfurt example that is used in a modal strategy against PAP(CC) need not eliminate every alternative possibility, but only robust alternative possibilities. One might think that even if Bribe fails as a modal counterexample to PAP (since *some* alternative possibilities remain available to the agent), it can be used as a modal counterexample to PAP(CC).

However, in the next section we shall consider a powerful objection to using prior sign cases like Bribe in a modal strategy against PAP(CC).

Even if it turns out to be impossible to provide a modal counterexample to PAP(CC), there is another strategy available. Frankfurt defenders might show that PAP(CC) is false by demonstrating that the availability of robust alternatives is not part of the conceptual fabric or metaphysical nature of moral responsibility. As Felipe Leon and Neal Tognazzini (2010) put it, alternative possibilities and moral responsibility might be “companions” without being “soul mates.” That is, alternative possibilities might happen to show up whenever moral responsibility does even if there is no deeper conceptual connection between them. Alternative possibilities might “tag along” with instances of moral responsibility even if the availability of alternative possibilities in no way enters into what it is to be morally responsible. As Derk Pereboom explains,

[I]t may be that a condition is necessary for some phenomenon A but sometimes holds by virtue of features that do not illuminate A, features that are irrelevant to explaining the nature of A. I believe that even if it turns out that conditions of the sort at issue [i.e. the availability of alternative possibilities] are necessary for moral responsibility, they can hold by virtue of features of a situation that are explanatorily irrelevant to what would make an agent morally responsible, and that as a result the condition at issue fails to illuminate the nature of the phenomenon. (Pereboom, 2001, p. 25)

The Frankfurt defender might attack PAP(CC) by showing that robust alternatives merely “tag along” with moral responsibility. Call this the “non-modal strategy.”

Consider again the case above. Suppose someone challenged the claim that the tautology ( $P \vee \sim P$ ) is not part of the conceptual fabric of what it is to be morally

---

<sup>12</sup> I shall consider other arguments in support of PAP in Chapter 4.

responsible. How might we show that an agent is not morally responsible (partly) in virtue of  $(P \vee \sim P)$ ? It is not possible to provide a modal counterexample since  $(P \vee \sim P)$  is true in every possible world. What we might do instead is provide a scenario in which  $(P \vee \sim P)$  plays no role in explaining why an agent performs the action he does. In this way, the skeptic should come to see that this tautology is explanatorily irrelevant to the agent's moral responsibility; it has nothing to do with whether he is morally responsible or not. Likewise, the Frankfurt defender might offer a Frankfurt example to show that whether or not the ability to do otherwise is a necessary condition for an agent's being morally responsible, it is explanatorily irrelevant.

Taking the analogy a bit further, suppose we did not know that  $(P \vee \sim P)$  is true in every possible world. Discovering that  $(P \vee \sim P)$  is explanatorily irrelevant to whether someone is morally responsible should undermine any confidence we might initially have had that  $(P \vee \sim P)$  is a necessary condition for moral responsibility. Absent further arguments in favor of such a claim, we should remain skeptical of the claim. And this, I shall argue, is the position we find ourselves in regarding PAP.

### ***2.5. The Dilemma Defense***

One of the most powerful objections to prior sign Frankfurt examples—one that can be used to defend PAP(CC)—is what has come to be known as the “dilemma defense.” It was initially suggested by Kane in a footnote (1985, pp. 51, footnote 25) and later developed by Kane (1996, pp. 142-144), David Widerker (1995) and Carl Ginet (1996). A key feature of Frankfurt examples like Bribe is that they involve a prior sign that reliably indicates what an agent is going to do unless some outside force intervenes. The dilemma defense targets this feature of the scenarios. It challenges proponents of prior sign Frankfurt examples to explain the nature of the connection between the prior sign and the agent's action. How, it is asked,

does the prior sign *reliably* indicate what Jones is going to choose to do on his own? There are only two options: either the prior sign is causally sufficient for Jones to decide to accept the bribe, or it is not. Either way, it is argued, the Frankfurt example fails.

Suppose the prior sign (or something it is indicative of) is causally sufficient for Jones to decide to accept the bribe. It is argued that the Frankfurt example begs the question against incompatibilists. After all, incompatibilists maintain that moral responsibility is ruled out by causal determinism. So an incompatibilist would not, and should not, be expected to accept the claim that Jones is morally responsible for his choice. However, if we suppose that the prior sign is *not* causally sufficient for Jones to choose to accept the bribe—that is, if the correlation between the prior sign and Jones’s choice is weaker than causal determinism—then it is argued that the intervener could not know what Jones is going to do unless he actually intervenes. After all, even if the prior sign indicates that it is highly probable that Jones will choose to accept the bribe, it remains possible that Jones chooses to do something else. But then it seems that Jones has *robust* alternative possibilities available to him. So, although it is dialectically appropriate to claim that Jones is morally responsible in such a scenario, he has robust alternative possibilities available to him, and thus, the Frankfurt example fails. Of course, Black could intervene regardless of whether Jones displays the prior sign or not, thereby ensuring that Jones chooses to accept the bribe. But then it would be dialectically inappropriate to claim that Jones is morally responsible for his choice under those circumstances.

Two general strategies for responding to the dilemma defense have emerged. One is to abandon prior sign Frankfurt examples in favor of some new and improved example (e.g., Hunt, 2000; McKenna, 2003; Mele & Robb, 1998, 2003; Stump, 1996, 1999, 2003; Widerker, 2009). Not surprisingly, the objections to these new examples turn upon the

issues raised in the dilemma defense: either the example illicitly presupposes causal determinism or the agent has a robust alternative available that is capable of grounding his moral responsibility. The other general strategy for responding to the dilemma defense is to take one of the two horns head on. Some have targeted the deterministic horn, arguing that prior sign Frankfurt examples that assume causal determinism need not beg the question against incompatibilists (e.g., Fischer, 1999, 2006a, 2010; Haji & McKenna, 2004, 2006). Others have targeted the indeterministic horn, attempting to provide prior sign Frankfurt examples that explicitly assume causal indeterminism while still eliminating every robust alternative possibility (e.g., Hunt, 2005; Pereboom, 2001, 2005, 2008). For the remainder of this chapter I shall focus on responding to the deterministic horn of the dilemma defense. This, I think, is the most promising avenue for rejecting PAP(CC). In the next chapter I shall consider attempts to provide indeterministic Frankfurt examples.

## ***2.6. The Deterministic Horn***

The modal strategy against PAP(CC) will not be successful against the deterministic horn of the dilemma defense. This is because the modal strategy requires the Frankfurt defender to present a metaphysically possible situation in which an agent is morally responsible yet lacks robust alternatives. However, claiming that an agent is morally responsible in a causally deterministic scenario presupposes compatibilism, thereby begging the question against incompatibilists.

This leaves only the non-modal strategy against PAP(CC) available to someone who wishes to respond to the deterministic horn. On this strategy, Frankfurt examples are intended to show that alternative possibilities play no role in our evaluation of an agent's actions. If this can be done without claiming that the agent is morally responsible, then the

Frankfurt defender will have avoided begging the question. This is the approach that Fischer takes in responding to the dilemma defense.<sup>13</sup>

Applying Fischer's strategy to Bribe, we explicitly assume that causal determinism obtains. Jones's exhibiting the prior sign is (or is indicative of something that is) causally sufficient for him to decide to accept the bribe. And were he to fail to exhibit the prior sign, Black would intervene, causing Jones to decide to accept the bribe. As things actually happen, Jones exhibits the prior sign, decides to accept the bribe, and Black never intervenes.

Crucially, Fischer does not conclude that Jones is morally responsible for his decision. Instead, he concludes that the fact that Jones lacked alternative possibilities is explanatorily irrelevant to whether Jones is morally responsible or not. Given that Black eliminates the relevant alternative possibilities, yet never *actually* intervenes, it seems that Black (and the fact that he eliminates Jones's ability to do otherwise) has nothing to do with whether Jones is morally responsible or not. To put it another way, if it turns out that Jones is not morally responsible for his decision to accept the bribe, it is not because he couldn't have done otherwise. After all, Black plays no role in an explanation of why Jones did what he did. So if Jones is not morally responsible for his decision, it is for some other reason. This, Fischer says, is the moral of the Frankfurt stories.

It is important to note that this conclusion leaves unanswered the question of whether causal determinism is compatible with moral responsibility. It is possible that the presence of alternative possibilities is irrelevant to whether an agent is morally responsible even if causal determinism *itself* rules out moral responsibility. Jones might be absolved of moral responsibility because he is causally determined and not because he lacks alternative

---

<sup>13</sup> See Fischer (1999, 2002, 2006a, 2010).



possibilities. Thus, the assumption of causal determinism does not beg the question against incompatibilists on this strategy.<sup>14</sup>

### ***2.7. Trouble for Fischer's Non-modal Strategy***

Stuart Goetz (2005) raises a worry for Fischer's strategy. Goetz says that causal determinism alone is sufficient to eliminate alternative possibilities. And, he argues that if alternative possibilities are already ruled out by causal determinism in prior sign Frankfurt examples, then the counterfactual intervener has no role to play in eliminating those alternative possibilities. Thus, whether the counterfactual intervener is also explanatorily irrelevant to the agent's moral responsibility is quite beside the point. As Goetz sees it, a Frankfurt example that assumes causal determinism,

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, 2005, p. 85)

---

<sup>14</sup> I agree with Fischer when he says that Ishtiyaque Haji and Michael McKenna (2004, 2006) are mistaken in claiming that it is question-begging to offer Frankfurt examples that assume causal determinism in a dialectical context where Frankfurt defenders are attempting to persuade incompatibilists that alternative possibilities are not required for moral responsibility.

Goetz's point is that if causal determinism is what really does the work of eliminating an agent's alternative possibilities, then the counterfactual intervener is explanatorily irrelevant to the elimination of an agent's alternative possibilities. But if this is the case, then for a Frankfurt defender to present an argument that hinges on the irrelevance of the presence of a counterfactual intervener to what the agent *actually* does misses the point. The counterfactual intervener is *already* irrelevant given the presence of causal determinism in the *actual* sequence of events. Therefore, Goetz argues, prior sign Frankfurt examples that assume causal determinism fail to demonstrate that the ability to do otherwise is explanatorily irrelevant to an agent's moral responsibility. That is, they fail establish the falsity of PAP(CC).

One way of responding to Goetz's objection is to claim that the counterfactual intervener itself is sufficient to ensure that the agent is unable to choose otherwise. This is the approach that Fischer (2006a) initially takes. He says that Jones's inability to choose otherwise is overdetermined. Causal determinism and the presence of Black are individually sufficient to eliminate the possibility that Jones chooses otherwise. Therefore, the counterfactual intervener *does* have a role to play. Fischer then argues that,

What the examples show is that *the mere fact* that Jones is unable to choose otherwise does not *in itself* establish that Jones is not morally responsible for his choice. This is because Black's counterfactual intervention is one of the factors that make it the case that Jones is unable to choose otherwise...and yet it is irrelevant to the grounding of Jones's moral responsibility. Considering this factor (the counterfactual intervention), and bracketing any other factor that might make it the case that Jones is unable to choose otherwise...it seems to me that Jones may well be morally responsible for his action. The mere fact that he lacks alternative possibilities, then, cannot in itself be the reason

that Jones is not morally responsible, if indeed he is not morally responsible. (Fischer, 2006a, pp. 199-200)

The mere presence of Black should be enough to say that Jones isn't morally responsible if the ability to do otherwise is really an essential part of what it is to be morally responsible. But it isn't since Black never activates his device. Were Black not there things would have gone the same way for Jones. Thus, the presence of Black plays no part in our considerations of whether Jones is morally responsible or not. And this suggests that the ability to do otherwise isn't really an essential part of what it is to be morally responsible.

The problem with this response, however, which Fischer (2010) acknowledges, is that the presence of Black alone is *not* sufficient for the elimination of Jones's ability to choose otherwise. Causal determinism must also be assumed. But if causal determinism *already* eliminates Jones's access to alternative possibilities, as Goetz claims it does, then Black has no role to play, and Goetz's objection prevails.

Fischer has revised his approach in response to Goetz's criticism. On this new approach, although we assume causal determinism, we also grant (at this stage of the argument) what Fischer calls the "agnostic assumption." That is, "*we take no stand* about whether causal determinism eliminates genuine metaphysical access to alternative possibilities" (Fischer, 2010, p. 325). In this situation causal determinism and Black's presence are taken to be jointly sufficient to rule out Jones's alternative possibilities, but we leave open the question of whether causal determinism is individually sufficient. Thus, at the outset, we cannot assume that Black has no role to play in the elimination of alternatives. From here, Fischer's argument proceeds almost exactly as before:

- (1) Black *plus* causal determinism rules out Jones's ability to do otherwise (and at this point we do not suppose that either would do so on its own).

- (2) Black is explanatorily irrelevant to whether Jones is morally responsible for deciding to accept the bribe.
- (3) The fact that Black rules out Jones's ability to do otherwise in a causally deterministic context is explanatorily irrelevant to Jones's moral responsibility.
- (4) Therefore, if Jones is not morally responsible for choosing to accept the bribe, it is not in virtue of the *mere* fact that he was not able to choose otherwise.<sup>15</sup>

Premise 2 is supported by considering the prior sign Frankfurt examples. Were Black removed from the situation, everything that has any causal influence on Jones would remain the same, and Jones's choice would be the same. Thus, if Jones is not morally responsible, it is not because of Black.<sup>16</sup>

What should we say about Fischer's strategy? I think it gets very close to succeeding. However, it doesn't work for the following reason. Someone sympathetic with Goetz's objection might respond to Fischer's argument in the following way:

In making the agnostic assumption, I grant that I am unable to claim that Black *is* irrelevant to the elimination of Jones's alternative possibilities, but it is still open to me

---

<sup>15</sup> This argument is adapted from Fischer (2010, pp. 328-330).

<sup>16</sup> Fischer (2013) has recently clarified the general structure of his argument in light of an objection from Widerker and Goetz (2013). They argue that Fischer's argument is an instance of the following invalid argument form:

- (1)  $p$  entails  $q$
- (2)  $p$  is (explanatorily) irrelevant to  $r$
- (3) Therefore,  $q$  is (explanatorily) irrelevant to  $r$ .

That this argument form is invalid can be seen by substituting 'Richard Nixon had to resign in 1974' for  $p$ , ' $2 \neq 1$ ' for  $q$ , and ' $2 > 1$ ' for  $r$ . However, Fischer says that the form of his argument differs from the above interpretation provided by Widerker and Goetz. His argument is of the following form:

- (1)  $p$  is one way of making it the case that  $q$
- (2)  $p$  is (explanatorily) irrelevant to  $r$
- (3) Therefore, the fact that  $q$  has been made true in some way is (explanatorily) irrelevant to  $r$ .

This argument form clearly avoids the counterexample posed by Widerker and Goetz.

to claim that, for all we know, Black *might* be irrelevant. After all, Black alone is not sufficient for the elimination of Jones's alternatives, and he might not be necessary—we can't say one way or another because of the agnostic assumption. And given this possibility, from the fact that Black is irrelevant to Jones's moral responsibility, we can't conclude that the *mere* lack of alternative possibilities is irrelevant to Jones's moral responsibility. After all, Black *might* have nothing to do with Jones's lack of alternative possibilities.

Given the agnostic assumption, we make no claim one way or another about whether causal determinism itself eliminates Jones's alternative possibilities. Moreover, Fischer grants that Black alone is not sufficient to rule out Jones's ability to do otherwise, but Black *plus* causal determinism is sufficient. This leaves us with two possible scenarios:

- (i) Causal determinism alone is sufficient to eliminate Jones's alternative possibilities and Black is not necessary.
- (ii) Causal determinism alone is not sufficient to eliminate Jones's alternative possibilities and Black is necessary.

Now, if we could rule out scenario (i), then Fischer's argument would go through. But we don't know which one is the case (at this stage of the argument). Given that both scenarios are (epistemically) possible, someone sympathetic with Goetz's objection cannot claim that Black *is* irrelevant, but he can say that Black *might* be irrelevant.

Fischer might respond by arguing that even if Black isn't necessary for the elimination of Jones's ability to choose otherwise, Black might still have an explanatory role to play. Later in his paper, once Fischer "relaxes" the agnostic assumption, he considers scenario (i) and the objection that Black is explanatorily irrelevant to the elimination of Jones's

alternative possibilities under those circumstances. Fischer argues that even in that scenario, we shouldn't conclude that Black is explanatorily irrelevant:

Why does the explanation in terms of causal determination “crowd out” all other explanations, including the explanation in which Black plays a crucial role?...Why is it *just obvious* that the existence of the one sort of explanation crowds out the other? Why, more specifically, is it *just obvious* that a prior state of the universe (together with the laws of nature) explaining why Jones cannot...choose otherwise leaves no room for *any other explanation* of Jones's inability—such as the presence of Black, his device, and his dispositions (in a causally deterministic context)? (Fischer, 2010, p. 332)

Fischer makes an excellent point. It might be that Black has an explanatory role to play, even in scenario (i). However, Fischer must do more than show that, for all we know, Black *could* play an explanatory role. This Goetz-inspired objection grants that. Rather, Fischer must show that Black *does* play an explanatory role. And it is not clear that Black does.<sup>17</sup>

---

<sup>17</sup> Consider the following situation (c.f., Fischer, 2010, pp. 330-331). Suppose we give a patient two medicines, M1 and M2, to cure Disease D. We do not know whether M1 itself is sufficient to increase the body temperature. But we do know that the combination of M1 and M2 is sufficient to increase the body temperature. Moreover, we know that M2 itself is not sufficient. We also know that M2 is causally inefficacious in curing the disease. From the preceding, can we conclude that the increase in body temperature itself is causally inefficacious in curing the disease? I don't see how we could. After all, it remains possible that M1 alone is sufficient to increase the body temperature. We do not know whether M2 is necessary. Thus, it might be that M1 cures Disease D by increasing the body temperature. We are in no position to say whether increasing the body temperature is causally efficacious.

It seems to me that this situation is analogous to the one we find ourselves in concerning Fischer's strategy. Given the agnostic assumption, we do not know whether causal determinism itself is sufficient to eliminate alternative possibilities. But we do know that the combination of causal determinism and Black's device (and dispositions) is sufficient to eliminate alternatives. Moreover, we know that Black's device alone is not sufficient. We also know that Black's device is explanatorily irrelevant to whether Jones is morally responsible. Can we conclude that if Jones is not morally responsible it is not because he lacked alternative possibilities? I don't think so. It still remains possible that causal determinism itself eliminates Jones's alternatives. Jones's device *might* be irrelevant to the elimination of Jones's alternatives. And if it is, then the elimination of alternatives by causal determinism might be relevant to whether Jones is morally responsible.

Perhaps a deterministic Frankfurt example could be devised where it is clear that the counterfactual intervener has a role to play in eliminating the agent's robust alternative possibilities. But none have been provided so far. And I have a difficult time imagining what one would look like. This leads me to conclude that Fischer's response to the deterministic horn of the dilemma defense, close as it comes to success, falls just short. However, I think that Fischer's general strategy is a good one and that so-called "complete blockage" Frankfurt examples can succeed where prior sign Frankfurt examples fail.

### ***2.8. Complete Blockage Frankfurt Examples***

Consider this scenario provided by John Locke:

[S]uppose a man be carried, whilst fast asleep, into a room where is a person he longs to see and speak with; and be there locked fast in, beyond his power to get out: he awakes, and is glad to find himself in so desirable company, which he stays willingly in...and yet, being locked fast in, it is evident he is not at liberty not to stay, he has not freedom to be gone. (Locke, 1979, p. 238)

This man does not leave the room, and he could not leave the room. Yet it appears that he is responsible for staying in the room. And were we to include a moral component to the situation—perhaps he knows that his child is in danger and is crying for help outside the door—we might say that he is morally responsible for staying in the room.

Notice that Locke's scenario does not make use of prior signs or counterfactual intervention. There is no mechanism that locks the door if the man is going to attempt to leave. Rather, the door is simply locked. His access to alternatives in which he leaves the room is blocked. However, we might object that the man is not morally responsible for failing to leave, but rather for not *trying* or *attempting* to leave, an action he did have an alternative to. So this case will not succeed against PAP(CC).

David Hunt (2000, 2003, 2005) offers a strategy for designing Frankfurt examples that are like Locke's locked room case, except that the walls of the room are "brought right inside the man's mind" (2005, p. 131). That is, the scenario is such that the man is prevented from even trying to do otherwise. In the blockage version of Bribe, Jones decides on his own to accept the bribe, but coincidentally, Black's device has blocked every neural pathway except the one Jones is actually using. Since Jones decides on his own, Hunt says that he is arguably morally responsible. Let us call such scenarios "complete blockage" cases.

Pereboom (2001) has provided an excellent illustration of the sort of case Hunt has in mind:

Imagine a universe correctly described by Epicurean physics: At the most fundamental level all that exists is atoms and the frictionless void, and there is a determinate downward direction in which all atoms naturally fall—except if they undergo uncaused swerves.

Situation C. A spherical atom is falling downward through space, with a certain velocity and acceleration. Its actual causal history is indeterministic because at any time the atom can be subject to an uncaused swerve. Suppose that the atom can swerve in any direction other than upwards. In actual fact, from  $t_1$  to  $t_2$  it does not swerve.

A counterfactual situation diverges from C only by virtue of a device that eliminates alternative possibilities and all differences thereby entailed:

Situation D. The case is identical to C, except that the atom is falling downward through a straight and vertically oriented tube whose interior surface is made of frictionless material, and whose interior is precisely wide enough to accommodate the atom. The atom would not have swerved during this time interval, and the trajectory,



velocity, and acceleration of the atom from  $t_1$  to  $t_2$  are precisely what they are in C.  
(Pereboom, 2001, p. 17)

In situation D the atom couldn't possibly move any differently than it does. The tube of frictionless material hedges the atom in so tightly that the atom couldn't even "bump" up against the tube. Yet, what the atom does is what it would have done whether the tube were present or not. Likewise, in the blockage version of Bribe, Jones is prevented from doing anything differently from what he in fact does. But had Black's mechanism not been present, Jones would have behaved in the exact same way, for the exact same reasons.

As Hunt sees it, the main challenge for complete blockage cases is in explaining, in a non-*ad hoc* way, how it is that they do not presuppose causal determinism. If a causally deterministic process blocks every alternative except Jones's choice to accept the bribe, then doesn't this amount to causal determinism? Robert Kane thinks so. He says,

[O]f course, there *are* no alternative possibilities left to the agent; every one is blocked except the agent's choosing [to accept the bribe at the exact moment he does]. But now we seem to have determinism pure and simple. By implanting the mechanism in this fashion, a controller would have predetermined exactly what the agent would do (and when); and, as a consequence, the controller, not the agent, would be ultimately responsible for the outcome. (Kane, 2000, p. 162)

Laura Ekstrom (2002) has also raised this worry. In a situation where Black's device blocks every alternative possibility, how could it be true, she asks, that Jones could never have done otherwise than he did yet he did everything on his own? "To say that [Jones] did something "on his own" is to say that he did it as the outcome of an unmanipulated, indeterministic deliberative process...An indeterministic, or nonnecessitating, cause is one that can fail to produce its effect, even without the intervention of anything to frustrate it" (Ekstrom, 2002,

p. 316). Provided that there is no event that makes Jones incapable of action or decision (e.g., his sudden death), the probability that Jones will decide to accept the bribe is equal to 1. For any event that occurs in the deliberative process, there was no chance that it would not occur given some prior event and the circumstances. Thus, Ekstrom concludes, it turns out that Jones's deliberative process "is a sequence of events connected entirely by deterministic causal relations" (Ekstrom, 2002, p. 317).

I am not convinced that complete blockage amounts to causal determinism.<sup>18</sup> But I shall not pursue the issue here. Let us grant that the complete blockage version of Bribe involves causal determinism. In granting this claim, it would certainly beg the question at issue between compatibilists and incompatibilists to also claim that Jones is morally responsible for his decision to accept the bribe. However, rather than deploy this case in the modal strategy, let us make use of Fischer's alternative approach. The conclusion we should draw from the complete blockage version of Bribe is that if Jones is not morally responsible for his choice, it is not because he lacks the ability to do otherwise. That is, Jones's lack of alternative possibilities is explanatorily irrelevant to whether he is morally responsible.

The argument proceeds as follows. Assume that Black's device blocks every neural pathway except the one Jones is actually using, hedging in Jones's deliberative process. As things actually happen, Jones decides to accept the bribe, and, coincidentally, his deliberative process proceeds exactly as it would have had Black's device not been activated. That is, Black's device, which eliminates Jones's alternatives, has nothing to do with what Jones does and why he does it. Jones does not decide to evade taxes *because* Black's device eliminated his alternatives. He decides as he does for his own reasons and

---

<sup>18</sup> For brief attempts at drawing a distinction between complete blockage and causal determinism, see Hunt (2003, p. 173), Mele and Robb (2003, p. 130), and Pereboom (2012, p. 305).

the device just happens to hedge in the corresponding neural pathway. Thus, it seems that the fact that Black's device rules out Jones's ability to do otherwise is irrelevant to an assessment of his moral responsibility for his decision. As Frankfurt says,

When a fact is in this way irrelevant to the problem of accounting for a person's action it seems quite gratuitous to assign it any weight in the assessment of his moral responsibility. Why should the fact be considered in reaching a moral judgment concerning the person when it does not help in any way to understand...what made him act as he did? (Frankfurt, 1969, p. 837)

Since the fact that Black's device rules out Jones's ability to do otherwise is irrelevant to an explanation of what he did and why he did it, it is difficult to see why it should have any bearing on Jones's moral responsibility.<sup>19</sup> The elimination of his alternatives plays no part in explaining why he did what he did.<sup>20</sup> Therefore, we have good reason to reject PAP(CC). PAP(CC) says that what it is to be morally responsible for performing some action is (in part) to have access to alternatives to performing that action. But there appear to be possible situations in which whether an agent has access to alternative possibilities has nothing to do

---

<sup>19</sup> The general principle I am employing here is the following:

Necessarily, if a fact is irrelevant to a full explanation of what action an agent performed and why he performed it, then such a fact has no bearing on the agent's moral responsibility for the act.

I shall have more to say about this principle in Chapter 4 when I consider several objections from David Widerker (2003) to this line of reasoning.

<sup>20</sup> The argument may be structured to parallel Fischer's (2010) argument:

(1') Black's device rules out Jones's ability to do otherwise.

(2') Black's device is explanatorily irrelevant to whether Jones is morally responsible for deciding to accept the bribe.

(3') The fact that Black's device rules out Jones's ability to do otherwise is explanatorily irrelevant to Jones's moral responsibility.

(4') Therefore, if Jones is not morally responsible for choosing to accept the bribe, it is not in virtue of the *mere* fact that he was not able to choose otherwise.

with whether the agent is morally responsible. The fact that Jones lacks alternative possibilities does not *in itself* explain why Jones is not morally responsible (if in fact he is not). But if PAP(CC) were true, it would.

I find this a compelling argument for rejecting PAP(CC). It does not beg the question even if we grant Kane and Ekstrom's claim that complete blockage amounts to causal determinism since I do not claim that Jones is morally responsible. Rather, I claim that the elimination of every alternative possibility does not *in itself* rule out moral responsibility. It may very well be that Jones is *not* morally responsible. What I am claiming is that the reason for this fact (if it is a fact) is not to be found in the *mere* elimination of alternative possibilities.

But we can go further. Given that the elimination of every alternative possibility amounts to causal determinism, and given that the complete blockage case shows that it is possible that the elimination of every alternative possibility is explanatorily irrelevant to an agent's moral responsibility, we can conclude that it is possible that causal determinism is explanatorily irrelevant to an agent's moral responsibility. That is, the mere fact of causal determinism does not *in itself* rule out moral responsibility. If an agent is not morally responsible, simply pointing out that he was causally determined will not be enough to account for this fact. His being causally determined might have had nothing to do with what he did or why he did it. Now, it still might turn out that causal determinism is incompatible with moral responsibility. Perhaps there is some *other* factor that both grounds causal determinism and is itself incompatible with moral responsibility (or perhaps causal indeterminism and moral responsibility just happen to be "companions"). But what this blockage case shows is that if we grant Kane and Ekstrom's claim, the reason for accepting the incompatibilist thesis is not to be found in causal determinism itself nor in the

elimination of alternative possibilities by causal determinism. Therefore, incompatibilists should not reject the conclusion of this argument simply because they are already committed to the incompatibility of moral responsibility with causal determinism.

One might resist this last step by giving up the claim that the elimination of every alternative possibility amounts to causal determinism. However, in making this move, it would no longer be question-begging to say that Jones could be morally responsible for his decision. But then it is open to the opponent of PAP(CC) to deploy the complete blockage version of Bribe as a modal counterexample to PAP(CC).

### ***2.9. Concluding Remarks***

What the complete blockage Frankfurt example shows is that it is possible that something that eliminates an agent's alternatives is irrelevant to an explanation of what the agent did and why he did it. After all, it could be that removing the device that makes it impossible for an agent to do otherwise would not affect what happened or why it happened in any way. But if this is so, then it is possible that the presence of the device is explanatorily irrelevant to the agent's moral responsibility. Since the device, which eliminates an agent's alternatives, could be explanatorily irrelevant to the agent's moral responsibility, it follows that the fact that the device rules out the agent's ability to do otherwise could be explanatorily irrelevant to his moral responsibility. But if PAP(CC) is true—an agent is morally responsible for his actions (partly) in virtue of his ability to do otherwise—then the fact that the device rules out this ability to do otherwise could not be explanatorily irrelevant to whether he is morally responsible. Therefore, PAP(CC) is false.

In using Frankfurt examples to challenge principles like PAP(CC), Frankfurt defenders have predominantly attempted to provide scenarios in which circumstances eliminate all of an agent's robust alternatives without eliminating every alternative possibility. This is

because of the worry that the elimination of every alternative amounts to causal determinism, thereby begging the question against incompatibilists. However, as we have seen, even if we grant that the elimination of every alternative does amount to causal determinism, this in itself is no strike against the Frankfurt example, provided that it is used in a non-modal strategy against PAP(CC).

I have argued that the primary motivation for accepting PAP is found in PAP(CC), and I have argued that complete blockage Frankfurt examples can be utilized as non-modal counterexamples to PAP(CC). If I am correct, then absent any other arguments in favor of PAP, we should be skeptical of its truth.<sup>21</sup> This is because there is no deep conceptual connection between moral responsibility and the ability to do otherwise. PAP *could* be true for all we know, but without this connection, I see little reason to believe it is. In the next chapter I shall consider several attempts at providing indeterministic Frankfurt examples that can be used in a modal strategy against PAP(CC). I will argue that, unfortunately, they all fail.

---

<sup>21</sup> Of course, there are other arguments in favor of PAP. See, for example, Widerker (2000), Copp (2003), and Zimmerman (2003). These arguments will need to be dealt with if we are to successfully undermine PAP. I shall turn to these arguments in Chapter 4.

### **Chapter 3. Indeterministic Frankfurt Examples**

The most significant challenge to prior sign Frankfurt examples is the dilemma defense. Recall that prior sign examples involve some sign, *F*, that indicates what the agent is going to do; if the agent exhibits *F*, then he will perform the action on his own and the counterfactual intervener need not interfere; if the agent does not exhibit *F*, then the counterfactual intervener must intervene to ensure the outcome he desires. The worry is that the only way the presence of *F* could ensure that the agent performs the action in question is if *F* (or something *F* is indicative of) is causally sufficient for the agent to perform the action. But this begs the question against incompatibilists. On the other hand, if *F* is not causally sufficient for the agent to perform the action in question, then the agent seems to have a robust alternative.

In response to the dilemma defense, some Frankfurt defenders have attempted to provide Frankfurt examples that do not rely on a prior sign. In the last chapter I considered one of these: complete blockage. Complete blockage cases ensure that the agent performs a particular action by eliminating every alternative available to an agent. A key objection to complete blockage cases is that the elimination of every alternative possibility amounts to causal determinism. Rather than resist this claim, I used the example in a non-modal argument against PAP(CC) that sidesteps the worry. Some, however, might want something more decisive, a modal counterexample to PAP(CC). In order to provide one that does not beg the question against incompatibilists, a Frankfurt example must be given that does not assume causal determinism (or what amounts to it). In this chapter I shall consider four leading approaches to providing indeterministic Frankfurt examples that could be used in the modal strategy. Unfortunately, all of them face significant problems.

### ***3.1. Occurrent Preemption Examples***

In “Rescuing Frankfurt-style Cases”, Alfred Mele and David Robb (1998) offer an indeterministic Frankfurt example that does not require the use of a prior sign:

Our scenario features an agent, Bob, who inhabits a world at which determinism is false...At  $t_1$ , Black initiates a certain deterministic process  $P$  in Bob’s brain with the intention of thereby causing Bob to decide at  $t_2$  (an hour later, say) to steal Ann’s car. The process, which is screened off from Bob’s consciousness, will deterministically culminate in Bob’s deciding at  $t_2$  to steal Ann’s car unless he decides on his own at  $t_2$  to steal it or is incapable at  $t_2$  of making a decision (because, for example, he is dead by  $t_2$ )...The process is in no way sensitive to any ‘sign’ of what Bob will decide. As it happens, at  $t_2$  Bob decides on his own to steal the car, on the basis of his own indeterministic deliberation about whether to steal it, and his decision has no deterministic cause. But if he had not just then decided on his own to steal it,  $P$  would have deterministically issued, at  $t_2$ , in his deciding to steal it. Rest assured that  $P$  in no way influences the indeterministic decision-making process that actually issues in Bob’s decision. (Mele & Robb, 1998, pp. 101-102, footnotes omitted)

Mele and Robb conclude that, all things being equal (e.g., Bob is sane), Bob is morally responsible for his choice to steal Ann’s car at  $t_2$  even though he was unable to avoid making that choice at  $t_2$ . Why think Bob is morally responsible? Because he decided to steal Ann’s car “on his own,” without any interference from  $P$ . Moreover, claiming that Bob is morally responsible in such a case does not beg the question against incompatibilism since Jones’s decision is not causally determined by some prior event.

An essential feature of Mele and Robb’s Frankfurt example is that the indeterministic deliberative process (call it ‘ $x$ ’) blocks (or preempts) the deterministic process  $P$  from



causing Bob to decide to steal Ann's car at  $t_2$ . Were this not the case,  $P$  would have a causal role to play in Bob's decision at  $t_2$ , and it would not be true that Bob decides "on his own."

Mele and Robb say that, as things actually happen,

*P* is preempted by  $x$  at  $t_2$ . *P* is not sensitive to a prior sign of Bob's decision. At every moment up to  $t_2$ , *P* is poised to cause Bob to decide at  $t_2$  to steal Ann's car. *P* is deactivated (preempted) only at the very moment that Bob makes his decision. (Mele & Robb, 2003, p. 129)

However, one might doubt that this is possible.<sup>1</sup> How could it be true that both  $x$  preempts  $P$  at  $t_2$  *and* had  $x$  not preempted  $P$  at  $t_2$ ,  $P$  would have caused Bob to decide to steal Ann's car at  $t_2$ ? One might worry that this case runs afoul of the indeterministic horn of the dilemma defense. Since the connection between  $x$  and Bob's decision to steal Ann's car is undetermined,  $P$  must wait to see whether  $x$  causes Bob to decide to steal the car at  $t_2$ . But if turns out that  $x$  does not cause Bob to steal the car at  $t_2$ , then it is too late for  $P$  to cause Bob to decide to steal Ann's car at  $t_2$ ; the moment has already passed. If this is correct, then despite Mele and Robb's claim to the contrary, it seems that Bob could have done otherwise at  $t_2$ .

In anticipation of this worry, Mele and Robb appeal to what Douglas Ehring calls "occurrent preemption."<sup>2</sup> Rather than blocking some intermediate event within process  $P$ ,

---

<sup>1</sup> See Widerker (2000), Timothy O'Connor (2000), and Ginet (2003; 2010).

<sup>2</sup> Mele and Robb note the difference between what they call "occurrent preemption" and what Ehring calls "occurrent preemption:"

Ehring defines preemption in terms of 'the same effect' (p. 20), and we see no need to insist that a decision to steal Ann's car that is produced by  $x$  at  $t_2$  is the *same* event or effect as the decision to steal Ann's car that is produced by  $P$  at  $t_2$  in the counterfactual scenario...we are not concerned to block nonrobust alternative possibilities or what John Fischer has called 'flickers of freedom' (1994, ch. 7). For example, it is fine with us that Bob has the following alternative possibilities: deciding to steal Ann's car as an upshot of his own indeterministic deliberation; and deciding to steal Ann's car as an upshot of

the blocking action allows P to run unimpeded but prevents the last event in P from bringing about the final effect. Ehring explains, “The efficacy of this last event is blocked, but not by preventing the occurrence of any intermediary events between that event and the final effect. Unlike in nonoccurrent preemption, the final event in the alternate line [i.e., P] would have *directly* caused the final effect (had it been efficacious at all)” (Ehring, 1997, p. 21).

Mele and Robb provide a model intended to show that occurrent preemption is possible in their Frankfurt example. We are to imagine a widget-making machine such that the color of the widget produced is determined by the color of a bb that hits the machine’s receptor. When a bb strikes the receptor and the machine is not already producing a widget, the machine immediately begins producing a widget. The machine only produces one widget at a time. Should two or more bbs hit the receptor simultaneously, the machine will produce a widget of the color of the right-most bb. In this way, say Mele and Robb, we have occurrent preemption: when multiple bbs strike the receptor simultaneously, the right-most bb causes the machine to produce a widget of a corresponding color and preempts the other bbs from causing the machine to produce a colored widget without interfering in the process that leads up to their striking the receptor. Likewise, in Bob’s case, should both neural processes P and x terminate at the same neural receptor, x’s terminating at the receptor causes Bob’s decision to steal Ann’s car and preempts P from causing Bob to make that decision without interfering with any of the events in P.

---

process P...Nor are we concerned that the actual and counterfactual deciding be the same event. What we call ‘occurrent preemption’ is consistent with this openness. (Mele & Robb, 2003, p. 134)

### 3.2. Criticism of Occurrent Preemption Examples

Some have objected that the use of occurrent preemption in Mele and Robb's Frankfurt example is incoherent.<sup>3</sup> I don't find these objections convincing.<sup>4</sup> But I won't pursue the matter further here. The more fundamental problem, as I see it, arises not from considering what happens when processes P and x converge at  $t_2$ , but rather, what happens if they diverge. As things actually happen, x terminates in the same neural receptor as P (call it 'N1'), culminating in a decision to steal Ann's car at  $t_2$ . But what if x had terminated in some other neural receptor (call it 'N2') culminating in a decision that is incompatible with the decision to steal Ann's car at  $t_2$  (e.g., a decision *not* to steal Ann's car)? The worry is that the possibility of divergence indicates that robust alternative possibilities are in fact available to Bob.

Mele and Robb respond to this worry by saying that "by  $t_2$  P has neutralized all of the [neural] nodes in Bob for decisions that are contrary to a decision at  $t_2$  to steal Ann's car" (Mele & Robb, 1998, p. 105). That is, P prevents the activation of any neural node (or its cognate), like N2, that would issue in some decision incompatible with the decision to steal Ann's car at  $t_2$ . P does not affect what happens *in* x, nor does it prevent x from terminating in one of these nodes. Rather, P prevents these nodes from activating, were x to terminate in one of them.

The trouble with this response is that in Mele and Robb's case, what Bob seems to be non-derivatively morally responsible for is not his decision to steal Ann's car at  $t_2$ , but rather, for the termination of his indeterministic deliberative process in N1 at  $t_2$ , which

---

<sup>3</sup> See Widerker (2000) and Ginet and Palmer (2010).

<sup>4</sup> See Kevin Timpe (2008, pp. 42-48) for several examples of occurrent preemption that could be of use to Mele and Robb.

culminates in his decision to steal Ann's car. But then it seems as though Bob did have a robust alternative. Although *x* actually terminates in *N1* at *t2*, it might not have. Bob's deliberative process could have terminated in *N2* at *t2*. And one might quite plausibly think that the availability of this alternative explains (in part) why Bob is morally responsible for his decision to steal the car at *t2*.<sup>5</sup> Process *x* is like the man in Locke's locked room case from Chapter 2 who was unable to leave, but could have attempted to. Although *x* is incapable of activating some neural node that issues in a competing decision, it could still terminate in one. What the case needs is a way of blocking *x* from terminating in one of these neutralized neural nodes. That is, it needs something like David Hunt's complete blockage. But once we introduce complete blockage into the case, we are in no better position than we were in Chapter 2.

### ***3.3. Limited Blockage Examples***

Michael McKenna (2003) says that the trouble with most attempts at designing Frankfurt examples is that in eliminating all of the alternatives within an agent's voluntary control, they come "dangerously close to making that problematic deterministic assumption" (McKenna, 2003, p. 206). But, an attempt to avoid this problem by relaxing the circumstances to make more room for alternatives "seems to invite sufficient slippage that the incompatibilist will be able to locate some robust alternative" (McKenna, 2003, p. 206). McKenna is confident that he can navigate his way between these difficulties. His approach is to provide a Frankfurt example where the only alternatives available to the agent within his voluntary control are non-robust alternatives—"alternatives that could *not* aid in grounding the judgment that an agent in a Frankfurt example is morally responsible for what

---

<sup>5</sup> See Robinson (2012) for why Fischer's (1994) response to the flicker defense cannot be successfully applied to Frankfurt examples such as Mele and Robb's.

she does” (2003, p. 206). Specifically, McKenna aims to eliminate every deliberatively significant alternative from the Frankfurt example, leaving only deliberatively *insignificant* alternatives within an agent’s voluntary control—alternatives in virtue of which the agent could not be held morally responsible.

What is the difference between deliberatively significant and deliberatively insignificant alternatives? Thinking it unlikely that a general principle can be offered to distinguish them, McKenna offers several cases to illustrate the distinction. Here is one of them, Needed Medication:

Tal arrives at Daphne’s house and discovers Daphne unconscious and in immediate need of a drug known as *The Good Stuff*. Unknown to Tal, Daphne has stored The Good Stuff in an aspirin jar. Grant that it is actionally open to Tal to walk to the medicine cabinet and retrieve for Daphne, from the jar marked ‘aspirin,’ The Good Stuff that Daphne needs to survive. It cannot be morally expected of Tal that he consider the option of fetching The Good Stuff from the aspirin jar even though this option is not causally closed to Tal. This mundane epistemic constraint on moral deliberation surely suggests that the range of morally significant options relevant to evaluating an agent’s responsibility should be restricted in some manner. It cannot be a demand upon moral agents that they possess the epistemic powers of deities or the wisdom of saints. (McKenna, 2003, p. 208).

There is simply no reason available to Tal to give Daphne whatever is stored in the jar marked ‘aspirin.’ Thus, although it is within Tal’s voluntary control to give Daphne The Good Stuff, such an option is not deliberatively significant. It is not the sort of alternative that it would have been reasonable for Tal to have considered as an alternative, given her circumstances.

With this distinction in mind, McKenna offers what he calls a “limited blockage”

Frankfurt example, Brain Malfunction:

The mild-mannered philosophy professor Casper comes upon a completely unexpected and highly unusual opportunity. He has just entered a room and is standing in front of a technologically state of the art ‘Make-it-the-Case Device’. Assume that Casper is justified in his true belief that the Make-it-the-Case Device is reliable and not merely a hoax.

On a large television screen at the top of the Make-it-the-Case Device appears a man dressed as a genie. The genie speaks:

Casper, just beneath this screen are two buttons, one marked ‘The Morally Good Thing to Do’ and another marked ‘The Morally Bad Thing to Do’. Let us abbreviate them as ‘Good’ and ‘Bad’ respectively. If you press the Bad button you will immediately make it the case that one million dollars are deposited into your bank account. The money will be drawn, in one-dollar increments, from the savings accounts of one million college professors. The transaction will be untraceable. If you press the Good button you will immediately make it the case that an entire village of people in the Amazon is cured of an otherwise fatal disease. Saving the villagers will not involve any money and by doing so you will not be stealing from your peers. You cannot select both buttons and this opportunity will not present itself again. You have ten seconds to select your option.

A timer appears on the screen and begins to count down from ten.

Casper pauses to consider these two options, quickly assessing the import of each...Imagining that shiny red Mercedes convertible roadster in the window, as the counter ticks away from 3 seconds to 2 greedily he takes the plunge and presses the Bad button. ‘Ah, dinner out tonight!’ Casper thinks to himself.

As it turns out, although Casper was unaware of this difficulty, and although there is no reason Casper should have been aware of this difficulty, at the time at which Casper greedily decided to press the Bad button, Casper had a small lesion on his brain that blocked the neural pathway constitutive of (or correlated with) a decision to push the Good button during that ten second interval. Casper could not have decided to press the Good button. (McKenna, 2003, pp. 209-210)

Assuming that Casper exercises libertarian free will (of whichever sort one prefers) and that the brain lesion played no part in the reasons that led Casper to press the Bad button, McKenna says that we have a metaphysically possible situation in which Casper is morally responsible for his decision to press the Bad button despite the fact that the only alternatives available to Casper were incapable of grounding his moral responsibility. McKenna points out that although Casper had “oodles and oodles” of alternatives, none of them were deliberatively significant. He could have “sung a little ditty and done a cutesy jig like Shirley Temple, finishing off with a set of jazz hands; or begin citing nursery rhymes; or made an attempt to eat his fist; or any number of equally ludicrous and irrelevant things” (McKenna, 2003, p. 213). The only thing Casper is prevented from doing during the interval of time is deciding to press the Good button. He could have done anything else, including *not deciding* to press the Bad button and *deciding not* to press the Bad button.

### ***3.4. Criticism of Limited Blockage Examples***

One might argue that deciding not to press the Bad button is a deliberatively significant alternative available to Casper. After all, he should be aware of the fact that by deciding not to press the Bad button he is not stealing from his peers. However, McKenna thinks that the scenario can be refined such that this alternative is not deliberatively significant for Casper.

We can imagine that this option was to Casper as...looking for the Good Stuff in the aspirin jar was to Tal in the example Needed Medicine. Knowing something about Casper's general framework of values, we might simply build into the case that Casper would find this option irrelevant. We might reason as follows: why in the world would such an alternative have occurred to him as a morally significant one? From his perspective he can either steal one million dollars from his peers in minor increments, or he can save a village from a deadly disease. Why would it occur to him to *not* steal a petty sum from each of one million of his peers, but to forgo saving the lives of the villagers? For Casper, it would be a perverse appreciation of the weight of moral reasons if Casper found the former compelling but not the latter. (McKenna, 2003, p. 211)

The idea is that from Casper's perspective he is not simply deciding between either stealing from his peers *or* not stealing from his peers. Rather, he is deciding between either stealing from his peers and failing to save a village from a deadly disease *or* saving a village from a deadly disease and not stealing from his peers. Given that Casper is unaware of the fact that he is incapable of choosing to press the Good button, it is difficult to see why the option of failing to press either button should occur to Casper as a relevant alternative. If he is not going to steal from his peers, then why not also cure the villagers in the Amazon?

Michael Robinson (2014) has recently objected to this reply. He points out that McKenna clearly rejects the notion that "the scope of reasonableness from an agent-relative perspective is limited to the values and standards endorsed by the agent" (2003, p. 209). What one finds reasonable might differ from what one *ought* to find reasonable. "What the white supremacist finds reasonable," says McKenna, "differs from what one ought to find reasonable" (2003, p. 209). But it is what one ought to find reasonable that matters when we



are considering whether an alternative is reasonable for an agent to consider at a particular time. Robinson argues that the decision not to press the Bad button is an alternative that Casper *ought* to find reasonable, regardless of whether he in fact finds it reasonable. Therefore, by McKenna's own lights, it is not enough for him to simply stipulate that Casper doesn't find the alternative reasonable, he needs to demonstrate that it is not the case that Casper ought to find it reasonable. But Robinson thinks that McKenna cannot meet this requirement without undermining the case. He says,

[D]eciding not to do that which one knows to be wrong is always deliberatively significant: It is reasonable for an agent to consider it, and it is reasonable for us to expect an agent to consider it, as an alternative to deciding to do that which he knows to be wrong. The only time it would be unreasonable to expect this of an agent would be when, owing to some sort of diminished capacity, the agent is genuinely incapable of appreciating the moral significance of these options—in which case the agent would fail to satisfy certain epistemic conditions required for morally responsible agency. Thus, provided that Casper is morally responsible for what he does (as we are to suppose), deciding not to steal one million dollars from his peers is a morally significant alternative that it would have been reasonable for Casper to have considered, whether or not he happened to think so. (Robinson, 2014, pp. 441-442)

The trouble with Robinson's objection is that it presupposes that Casper ought to consider his options regarding each button individually: he ought to decide whether or not to press the Bad button and then decide whether or not to press the Good button. But why think that? There is another approach to decision-making available to Casper that seems to be at least as reasonable. He might have looked at the big picture, considered the

consequences of each decision, and then made a decision accordingly. On this approach, the following three options are available to Casper to consider:

- (1) Choose both to cure the Amazonian villagers and *not* to steal from his peers.
- (2) Choose both to steal from his peers and *not* to cure the Amazonian villagers.
- (3) Choose both *not* to steal from his peers and *not* to cure the Amazonian villagers.

Assuming that Casper does not want the people of the Amazon village to die of a fatal disease, he has no reason to take option (3) seriously. In this way, option (3) is similar to the alternative available to Tal in which she gives Daphne whatever is in the jar marked ‘aspirin.’ Casper could decide not to press either button, but why would he do so?

However, if Casper decides not to steal from his peers, then option (3) is the only alternative available to him since the brain lesion (of which Casper is unaware) prevents him from choosing option (2). So if we assume that Casper makes a decision (an assumption we shall challenge in a moment), then on the big picture approach the only reasonable option for Casper to choose is option (2). Option (3) is deliberatively insignificant and option (1) is ruled out by the brain lesion.

What we have, then, are two different approaches to decision-making available to Casper. It is not obvious that Robinson’s discrete approach is any more reasonable than the big picture approach. In fact, given that a decision regarding one button may effect Casper’s decision regarding the other button, the big picture approach is arguably the more reasonable approach. Thus, given that it is not obvious that Casper ought to prefer the discrete approach to the big picture approach, McKenna might rule out the discrete approach by simply stipulating that Casper is a big picture kind of guy who would not deliberate over each button individually.

There is another response available to McKenna, one he briefly mentions but does not pursue. McKenna might grant that deciding not to press the Bad button is deliberately significant for Casper and simply eliminate the alternative altogether by introducing another brain lesion. That is, rather than alter the case to make this alternative deliberately insignificant for Casper, McKenna might modify the case so that the alternative is no longer under Casper's voluntary control.

Robinson considers another alternative available to Casper, one Robinson believes is robust. Casper could simply omit to decide to press the Bad button.<sup>6</sup> Robinson concedes that since Casper is unable to actively form an intention (i.e., decide) not to press the Bad button, "it may be that it is not possible for him to *intentionally* omit to decide to press the Bad button" (2014, p. 442). But this is not a problem, Robinson thinks, because "an omission need not be intentional in order to constitute a robust alternative. It must be within an agent's control, to be sure, but an omission can be within an agent's control (even direct voluntary control) without being intentional" (2014, p. 442).

The kind of omission Robinson has in mind corresponds with a class of action that some have called "nonintentional action."<sup>7</sup> He provides the following illustration. Imagine a runner who wears down the soles of her shoes while running in a marathon. Suppose that

---

<sup>6</sup> Robison argues that there is no reason to think that only actions can be robust alternatives. Agents can have at least as much control over their omissions as their actions. For example, suppose I am sitting at the bus stop waiting for my bus to arrive. Once the bus arrives, there are many paths that are genuinely available to me. One path involves me getting on the bus. However, there are other paths that don't involve me getting on the bus. I can decide not to get on the bus. But I could also simply omit to decide to get on the bus. "This too is within my direct voluntary control. Just as there is nothing else I need to do in order to decide (not) to get on the bus—I simply decide—there is nothing I need to do in order to omit to (decide to) get on the bus; I simply omit to (decide to) do so" (2014, pp. 439-440).

<sup>7</sup> See Mele and Moser (1994); Mele and Sverdlik (1996).

she is consciously aware that she is wearing down the soles of her shoes while she is running, but she has no intention to do so, nor any intention whose content includes wearing down her shoes (e.g., an intention to *try* to wear down the soles). Given that the runner is in control of her actions and is aware that while she is running she is wearing down the soles of her shoes, it would be inaccurate to claim that she unintentionally wears down the soles of her shoes. Wearing down her shoes is a side effect of an intentional action. However, it would also be a mistake to claim that the runner intentionally wears down the soles of her shoes. This is because, as Robinson puts it, “wearing down the soles of her shoes in no way figures into the content of any intention she has with respect to running the marathon” (2014, p. 443). So the action of wearing down the soles of her shoes is neither *intentional* nor *unintentional*, but rather, it is *nonintentional*.<sup>8</sup> The runner knowingly and voluntarily performs the action without intentionally performing it. Similarly, says Robinson, “one can knowingly and voluntarily omit to perform an action without intentionally omitting to perform it” (2014, p. 443). However, unlike the example of the runner, where the nonintentional action is not within the agent’s direct voluntary control, nonintentional omissions can be within an agent’s direct voluntary control. Robinson says that “there is nothing else an agent needs to do *in order* to knowingly and voluntarily omit to perform some action; he simply omits to do so” (2014, pp. 443, n. 414).

An example might be helpful here.<sup>9</sup> Suppose that while I am admiring a freshly painted object I notice a sign in front of it that says, “Wet Paint. Do Not Touch.” Once I see the sign, I might simply refrain from touching the object. Although there may be several things

---

<sup>8</sup> Mele and Sverdlik define nonintentional action in the following way: “insofar as an agent who is A-ing is neither aiming at A-ing nor trying to A, either as an end or as a means to (or constituent of) an end, she is not intentionally A-ing; insofar as an agent is A-ing knowingly and nonaccidentally, she is not unintentionally A-ing; and actions that are neither intentional nor unintentional are nonintentional” (1996, p. 274).

I do while refraining (e.g., I might twiddle my thumbs or put my hands in my pockets), I do not need to perform some action in order to prevent myself from touching the object. Suppose I had my arms at my sides before I read the sign. That need not be an action I was performing then, and having my arms at my sides need not be an action I perform after reading the sign. Upon reading the sign I might have no inclination whatsoever to touch the object. There might be no temptation, and, therefore, no need to actively keep my arms at my sides. Furthermore, I do not need to actively form an intention (i.e., decide) to refrain from touching the object. Upon reading the sign, the question of whether to touch the object may never enter my mind. I may simply refrain from touching it without making a decision not to touch it. Indeed, I may not form any intention whose content includes refraining from touching the object (e.g., an intention to *try* not to touch the object). Thus, my omission may be nonintentional. It would be strange, however, to claim that I do not knowingly and voluntarily omit to touch the object. Likewise, in the modified version of Brain Malfunction, Casper may knowingly and voluntarily omit to decide to press the Bad button even though this omission is not intentional.

Robinson argues that omitting to decide to press the Bad button is a robust alternative for Casper:

Casper had it within his control not to decide to press the Bad button, which he could ensure by knowingly and voluntarily omitting to decide to press it; omitting to decide to press the Bad button was morally less bad than deciding to press the Bad button; and it would have been reasonable for Casper to have considered omitting to decide to press the Bad button as an alternative to deciding to press it given his agent-relative deliberative circumstances. (2014, p. 443)

---

<sup>9</sup> The following example comes from Clarke (2014, pp. 14, 63).

Thus, Brain Malfunction (even when modified) fails to provide a possible scenario in which an agent is morally responsible yet lacks robust alternatives.

What should we say about Robinson's reply to McKenna? I think it is successful. Without the introduction of something that ensures that Casper will make a decision to press a button (which would ensure that he decides to press the Bad button since the brain lesion prevents him from deciding to press the Good button), it is open to Casper to simply refrain from deciding to press the Bad button.

McKenna seems to have only two options available to him in dealing with this alternative. First, he could further revise Brain Malfunction so that, although this alternative is available to Casper, it is not deliberatively significant for him. McKenna might stipulate that Casper is not the kind of guy that would allow time to run out before he makes a decision. Therefore, Casper would not omit to decide to press the Bad button. It is just not the sort of alternative he would find reasonable, given his circumstances. However, the problem with this reply is that omitting to press the Bad button is an alternative he *ought* to find reasonable, regardless of whether he in fact finds it reasonable. It is reasonable for Casper to consider omitting to press the Bad button as an alternative to doing what he knows to be wrong, and it is reasonable for us to expect him to consider it as an alternative.<sup>10</sup>

McKenna's second option is to eliminate the possibility that Casper knowingly and voluntarily omits to press the Bad button by altering the case. The only way to do this is by introducing some mechanism that will ensure that Casper decides to press the Bad button if he doesn't do so on his own. But now the case seems to require either causal determinism or some sort of prior sign. If causal determinism is required, then Brain Malfunction is not

---

<sup>10</sup> This objection does not run afoul of my earlier reply to Robinson since the alternative Casper ought to be considering is not a *decision*, but an *omission*.

really an indeterministic Frankfurt example. If a prior sign is required, then the case fares no better than other prior sign Frankfurt examples.

Despite the foregoing, one might remain skeptical of the claim that nonintentional omissions can be robust alternatives. It might be thought that an alternative must be either an action or intentional if it is to be robust. However, even if we grant this restriction on robust alternatives, there is another alternative available to Casper. Instead of deciding to press the Bad button, he could have continued deliberating until time ran out. He could have weighed the reasons for and against pressing each button and simply failed to make a decision within the allotted time. Had he done so, he would not have decided to press the Bad button, and therefore would not be morally responsible for pressing the Bad button. Given how important a decision Casper is making, this alternative would seem to be deliberatively significant. At any given time, if Casper hasn't reached a decision yet, the most reasonable thing for him to do is to continue trying to reach one.

It is important to notice that in continuing to deliberate until time runs out, Casper is not deciding not to make a decision about which button to press. And he is not deciding that he will deliberate until time runs out. On the contrary, Casper could be doing his reasonable best to reach a decision; yet, despite his best efforts, he might simply fail. Time might run out before he comes to a decision. Indeed, given that Casper could not decide to press the Good button because of the brain lesion, one might think that this is exactly what he should have done.

Any attempt to eliminate this alternative or to render it deliberatively insignificant faces the difficulties raised above. McKenna cannot simply stipulate that Casper is not an indecisive deliberator, and would therefore reach a decision, because Casper *ought* to find deliberating until time runs out a reasonable alternative to doing what he knows to be wrong,

regardless of whether he does in fact find it to be so. And any attempt to introduce some mechanism that ensures Casper decides to press the Bad button if he doesn't do so on his own will require introducing a prior sign. Therefore, Brain Malfunction seems to be in no better shape than other prior sign Frankfurt examples.

David Widerker (2006, 2009) has suggested an approach to dealing with this seemingly robust alternative.<sup>11</sup> He offers his own limited blockage case, Brain-Malfunction-W:

Jones is deliberating as to whether to keep the promise he made to his uncle to visit him in the hospital shortly before his uncle is about to undergo a critical operation. Jones is his uncle's only relative, and the visit is very important to the uncle. The reason for Jones's deliberating is that, on his way to the hospital, he (incidentally) met Mary—a woman with whom he was romantically involved in his distant past, and whom he has not seen since then. Mary, being eager to talk to Jones, invites him for a cup of coffee in a nearby restaurant. She explains that she is in town just for a couple of hours, and wishes to spend those hours with him. Jones is aware that if he accepts Mary's offer, he will not be able to make it to the hospital during visiting hours. Normally, one can avoid deciding as one does by deciding otherwise. But in our scenario Jones does not have that option, since shortly after beginning to deliberate, he undergoes a neurological change as a result of which one of the (neurological) causally necessary conditions for his deciding otherwise, a condition which we may call '*N*', does not obtain. It is also assumed that

---

<sup>11</sup> Although Widerker now finds some Frankfurt examples intuitively appealing, he also finds the W-defense of PAP intuitively plausible. He says, "As it often happens in philosophical inquiry, in our attempt to clarify a certain concept, we end up being confronted with two different and equally attractive accounts of it, with a genuine difficulty deciding between them. This at any rate is how I feel with respect to the debate under consideration. Hence, I shall not try to adjudicate between these accounts" (2009, pp. 90-91). He remains an incompatibilist, but he believes that arguments for the incompatibility of moral responsibility and causal determinism should not rely on PAP. I shall discuss the W-defense in Chapter 4.



this fact is unknown to Jones (who believes that he can decide to keep the promise), and that *N*'s absence does not affect his deliberation process. In the end, Jones decides on his own not to keep the promise, and spends the afternoon with Mary. (Widerker, 2009, pp. 89-90)

As described, Jones could deliberate indefinitely, thereby avoiding deciding to break his promise. So Widerker adds the following:

To complete the example, we need to eliminate all the *actionally accessible* alternatives to Jones's deciding not to keep the promise that are within his power to realize, or at any rate, all the *morally significant* alternatives. For example, there is the alternative of his continuing to deliberate at *t*, instead of deciding not to keep the promise. However, this alternative can be ruled out either by stipulation, or by assuming that in the scenario at hand Jones needs to make a decision right away, as otherwise he will miss the bus that could get him to the hospital on time. In the latter case, the act of continuing to deliberate is morally equivalent to the decision not to keep the promise, and hence does not count as a *morally significant alternative*. (Widerker, 2009, p. 90)

Widerker says that the worrisome alternative of deliberating indefinitely can be rendered non-robust in two ways: by stipulation (though he doesn't say exactly how this might be done) or by introducing a time limit.

An attempt to stipulate the alternative away will not work, for the reasons given above. Jones ought to find deliberating a reasonable alternative to doing what he knows to be wrong. Any attempt to alter the case so that Jones does not find it a reasonable alternative threatens to undermine the claim that he is morally responsible for his decision not to keep his promise.

Attempting to render the alternative non-robust by introducing a time limit is also problematic. Widerker claims that Jones's continuing to deliberate until time runs out is morally equivalent to the decision not to keep his promise. But why think this is so? Widerker does not say. Perhaps he thinks that in order for Jones to continue deliberating until time runs out he must decide not to make a decision about whether to keep his promise. But I see no reason to accept this claim. Jones might try to reach a decision and simply run out of time.<sup>12</sup> Given the neurological change that precludes Jones from deciding to keep his promise to his uncle, it is reasonable to think this is exactly what Jones ought to do. Thus, Jones appears to have a robust alternative to deciding not to keep his promise.

### ***3.5. Internal Sign Examples***

Eleonore Stump (1996, 1999, 2003) has offered a Frankfurt example that requires a sign, but unlike prior sign examples like Bribe, the sign does not occur prior to the agent's performing the action. Instead, the sign is part of a sequence of neural firings that, taken as a whole, corresponds to the agent's action. Here is her Frankfurt example:

Suppose that neurosurgeon Grey wants his patient Jones to vote for Republicans in the upcoming election. Grey has a neuroscope which lets him both observe and bring about neural firings which correlate with acts of will on Jones's part. Through his neuroscope, Grey ascertains that every time Jones wills to vote for Republican candidates, that act of

---

<sup>12</sup> A person need not decide not to perform some action in order to be responsible for failing to perform that action; he might simply decide to perform some other action. Consider this case offered by Randolph Clarke:

Suppose that while walking in the countryside you come to a fork in the path. You're aware that the path on the left is more pleasant, and you realize that should you take the path on the right your walk will be less enjoyable. Suppose that you nevertheless decide to take path on the right (perhaps believing that path shorter), and you then do so, aware that in so doing you aren't taking the left path. It seems that you needn't intend not to take the left path in order for it to be the case that you intentionally don't take (omit to take) that path. (Clarke, 2010, p. 165)

his will correlates with the completion of a sequence of neural firings in Jones's brain that always includes, near its beginning, the firing of neurons a, b, c (call this neural sequence 'R'). On the other hand, Jones's willing to vote for Democratic candidates is correlated with the completion of a different neural sequence that always includes, near its beginning, the firings of neurons x, y, z, none of which is the same as those in neural sequence R (call this neural sequence 'D'). For simplicity's sake, suppose that neither neural sequence R nor neural sequence D is also correlated with any further set of mental acts. Again for simplicity's sake, suppose that Jones's only relevant options are an act of will to vote for Republicans or an act of will to vote for Democrats.

Then Grey can tune his neuroscope accordingly. Whenever the neuroscope detects the firing of x, y, and z, the initial neurons of neural sequence D, the neuroscope immediately disrupts the neural sequence, so that it isn't brought to completion. The neuroscope then activates the coercive neurological mechanism which fires the neurons of the neural sequence R, thereby bringing it about that Jones wills to vote for Republicans. But if the neuroscope detects the firing of a, b, and c, the initial neurons in neural sequence R, which is correlated with the act of will to vote for Republicans, then the neuroscope does not interrupt that neural sequence. It doesn't activate the coercive neurological mechanism, and neural sequence R continues, culminating in Jones's willing to vote for Republicans, without Jones's being caused to will in this way by Grey.

And suppose that...Grey does not act to bring about neural sequence R, but that Jones wills to vote for Republicans without Grey's coercing him to do so. (Stump, 2003, pp. 140, footnotes omitted)

Stump adds that we may suppose that Jones's act is indeterministic.

Importantly, Stump says that were the device to interrupt a neural sequence, it would not be true to say that Jones engaged in the mental activity of *beginning to choose*. Jones would have no mental act. Why not? Stump assumes a one-many relation between a mental act and neural sequence.<sup>13</sup> She says,

When I suddenly recognize my daughter's face across a crowded room, that one mental act of recognition, which feels sudden, even instantaneous, to me, is correlated with many neural firings as information from the retina is sent through the optic nerve, relayed through the lateral geniculate nucleus of the thalamus, processed in various parts of the occipital cortex, which take account of figure, motion, orientation in space, and color, and then processed further in cortical association areas. Only when the whole sequence of neural firings is completed, do I have the mental act of recognizing my daughter. Whatever neural firings are correlated with an act of will or intellect, I take it that in this case, as in all others, the correlation between the mental act and the firing of the relevant neurons is a one-many relation. (Stump, 1999, p. 417)

According to Stump, no mental act is performed unless the firing of the entire neural sequence correlated with the mental act is completed:

If the neural sequence correlated with my recognizing my daughter's face across a crowded room is interrupted at the level of the thalamus, say, then I will have no mental act having to do with seeing her. I won't for example, think to myself, "For a moment there, I thought I saw my daughter, but now I'm not sure." I won't have a sensation of almost but not quite seeing her. I won't have a premonition that I was about to see her,

---

<sup>13</sup> Stump intentionally leaves vague the nature of the correlation between the mental act and the sequence of neural firings in order to avoid excluding any current theory of the relation between the mind and the brain except Cartesian dualism.

and then I mysteriously just don't see her. I will simply have no mental act regarding recognition of her at all. (Stump, 1999, pp. 417-418)

Thus, according to Stump, Jones is not causally determined to choose as he does, yet he could not have chosen differently (or *began* to choose differently), and he seems to be morally responsible for his decision.

### ***3.6. Criticism of Internal Sign Examples***

However, as David Widerker (2000) points out, if Jones's action is not causally determined, then he very well might have robust alternative possibilities available to him. Widerker reasons as follows. Consider a scenario like the one described by Stump but without the existence of a counterfactual intervener. Assume that Jones exercises libertarian free will (of whichever sort one prefers). Widerker says,

In that scenario, there would be no reason to think that Jones could not have decided otherwise. Now recall that, on Stump's view of decisions, once the neural firings a, b, c occur, Jones...is bound to decide to vote for a Republican candidate. This means that the only way in which Jones could have decided otherwise in the above scenario, is by having the power to bring about the non-occurrence of a, b, c; a power that he would have before the occurrence of a, b, c and not after that. But if he has that power in the said scenario (as he surely does), he must also have it in the scenario featuring Grey. That the latter scenario includes a potentially coercive neuroscope does not change this fact, since its coercive influence would come into play only after the possible occurrence of x, y, z, that is, at a time that is later than the occurrence of a, b, c. Hence, it does not affect Jones's power to bring about the nonoccurrence of a, b, c. (Widerker, 2000, p. 186)

Thus, in reply to Stump, Widerker says that Jones could have failed to bring about the occurrence of a, b, c by bringing about x, y, z. Had Jones done so, he would not be morally responsible for choosing to vote for a Republican candidate.

Stump replies to this objection. Drawing from the quote above, she thinks it is clear that Widerker supposes that Jones's bringing about the firing of neurons a, b, c is an act that is distinct from and prior to the act correlated with neural sequence R. And it is because Widerker thinks Jones is able to refrain from performing this antecedent act (i.e., bringing about the firing of a, b, c) that he thinks Jones is able to do otherwise. However, Stump argues that this way of thinking about the situation is problematic.

She argues as follows. Call the decision to vote Republican 'A1' and the act of bringing about the firing of a, b, c 'A2.' Now, A2 will also be correlated with a neural sequence, for instance, one that begins with the firings of neurons r, s, and t.<sup>14</sup> However, says Stump, "on Widerker's reasoning that Jones acts to bring about the firing of neurons a, b, and c, it apparently follows that Jones must act to bring about the firings of neurons r, s, and t too" (2003, p. 147). Call this act 'A3.' A3 itself will correlate with another neural firing, which Jones will have to act to bring about (call this act 'A4'), and so on, in an infinite regress. Stump concludes that Widerker's view seems to require that Jones complete an infinite series of acts in order to do anything.

The problem with Widerker's objection, says Stump, "lies in his supposing that a willer such as Jones brings it about that neurons a, b, and c fire and that Jones's doing so is itself an act of Jones's antecedent to and separate from Jones's act of will to vote Republican" (2003, p. 147). However, I think Stump has missed the point. Widerker need not make this

---

<sup>14</sup> For this Frankfurt example Stump is targeting libertarians who are not Cartesian dualists. Later in her paper she offers another Frankfurt example that targets Cartesian dualists.

assumption. Given the assumption of indeterminism, it is open to Jones to bring it about that neurons a, b, c fire or that neurons x, y, z fire. Instead of bringing about the firing of neurons a, b, c, Jones could have brought about the firings of neurons x, y, z. Had he done so, Grey's neuroscope would have interrupted neural sequence D, then caused the neurons of neural sequence R to fire, and Jones would not be morally responsible for deciding to vote Republican. If Stump were to deny that this is a robust alternative, then it would be reasonable for a libertarian to reject Stump's claim that the agent's action is causally undetermined.

### ***3.7. Buffer Cases***

Derk Pereboom (2001) and David Hunt (2005) have provided prior sign Frankfurt examples where the agent in question is purported to have access only to non-robust alternatives yet is morally responsible. Unlike the Frankfurt examples offered by McKenna and Widerker, these cases do not involve directly blocking an agent's access to robust alternatives. Instead, there is an intermediate action—a “buffer”—that the agent must perform before he has access to robust alternatives. Although performing the intermediate action is a necessary condition for having robust alternatives, it is not sufficient, and it is here that the counterfactual intervener is inserted. If the agent were to perform this intermediate action, the counterfactual intervener would immediately intercede preventing the agent from accessing robust alternatives. But as things actually happen, the agent never performs this intermediate action and the counterfactual intervener never interferes. Since the agent must cross the “buffer zone” before having access to robust alternatives, these Frankfurt examples have come to be known as “buffer cases.”

Over the years Pereboom has offered several different versions of his buffer case. Here is his most recent (and preferred) version, “Tax Evasion (3)”:<sup>15</sup>

Joe is considering claiming a tax deduction for the registration fee that he paid when he bought a house. He understands that claiming this deduction is illegal, but he believes that he probably won’t be caught, and that if he were, he would plead ignorance and likely do so successfully. Suppose he has a strong but not always overriding desire to advance his self-interest regardless of its cost to others and even if it involves illegal activity. He is aware and sensitive to the moral reasons not to evade taxes in the way he is contemplating, but in this case these reasons are overridden by self-interest. If he were to refrain from deciding to evade taxes, it would be for self-interested reasons. In fact, it is causally necessary for his failing to decide to evade taxes in this situation that he imagine, with a certain degree of vividness, being at least fairly severely punished for doing so, a mental state he can produce voluntarily. (He knows that his jurisdiction allows for a prison sentence for the sort of tax evasion he is contemplating.) However, Joe’s imagining in this way being punished is not causally sufficient for his failing to choose to evade taxes. If he were to do so, he could, exercising his libertarian free will,

---

<sup>15</sup> In an earlier case of Pereboom’s, Tax Evasion (2), the necessary condition is Joe’s reaching a certain level of attentiveness to his moral reasons. Were he to do so, Black’s device would activate, resulting in Joe’s deciding to cheat on his taxes. David Widerker objects that, in Tax Evasion (2), what Joe is directly blameworthy for is failing to attain the requisite level of attentiveness to his moral reasons, and he is only derivatively blameworthy for his decision to evade taxes. He says,

[A] problem with Pereboom’s example is that, in it, the agent is *derivatively* blameworthy for the decision he made, because he has not done his reasonable best (or has not made a reasonable effort) to avoid making it. He should have been more attentive to the moral reasons than he in fact was—something he could have done. And in the case, he would not be blameworthy for deciding to evade taxes, as then he would be forced by neuroscientists so to decide. If this is correct, then Pereboom’s example is a case of derivative culpability, and hence is irrelevant to PAP, which...concerns itself only with direct or non-derivative culpability. (Widerker, 2006, p. 173)



either decide to evade taxes or refrain from so deciding (without the intervener's device in place). However, to ensure that he will decide to evade taxes, a neuroscientist [the infamous Black] has, unbeknownst to him, implanted a device in his brain, which, were it to sense his imagining with a certain degree of vividness being at least fairly severely punished for evading taxes, would electronically stimulate the right neural centers so as to inevitably result in his making this choice. As it happens, Joe does not imagine in this way being punished, and he decides to take the illegal deduction while the device remains idle. (Pereboom, 2009, p. 117)

Joe's imagining the punishment he could receive with a certain level of vividness is the prior sign that initiates intervention from the device. Robust alternative possibilities—such as Joe deciding not to evade his taxes—are on the other side of this buffer, says Pereboom. Joe can't access these robust alternative possibilities without first exhibiting the prior sign. But if he exhibited the prior sign, if he imagined the punishment he could receive with a certain level of vividness, the device would intervene, ensuring that Joe decides to evade his taxes. Thus, according to Pereboom, although Joe has access to alternative possibilities under his control—he can vividly imagine his punishment or not—these alternatives are not robust. If this is right, then we seem to have a Frankfurt example that can be used as a modal counterexample to PAP(CC): Joe is morally responsible for his decision to evade his taxes even though he lacks robust alternative possibilities.<sup>16</sup>

Other prior sign Frankfurt examples, like Bribe in Chapter 2, also include a necessary condition for doing otherwise. Tax Evasion (3) diverges from those cases in one important respect: the absence of the necessary condition at any given time is not a causally sufficient

---

<sup>16</sup> To forestall the objection that Joe is not responsive to moral reasons, Pereboom has subsequently added that if Joe's moral reasons were stronger, Joe could decide not to evade taxes for moral reasons alone.

condition for the agent performing the action she does (Pereboom, 2009, p. 113). That is, in Tax Evasion (3), Joe's failure to vividly imagine his punishment to the requisite degree at any given moment is not a sufficient condition for Joe's deciding to evade taxes. As Pereboom says of the necessary condition, "Its absence at any specific time is compatible with both Joe's deciding to evade taxes and with his not deciding to evade taxes" (Pereboom, 2009). This is important because it allows it to be the case that at no time is Joe causally determined to evade taxes (see Pereboom, 2001, pp. 9-10, 20; 2009, p. 113; 2014, pp. 15-16; and Widerker, 1995, pp. 249-250). Were the absence of the necessary condition causally sufficient to evade taxes, we would simply have a prior sign case and would be back to square one with regard to the dilemma defense.

One might wonder why the possibility that Joe enters the buffer zone is not a robust alternative. After all, by voluntarily entering the buffer zone (i.e., vividly imagining the punishment he could receive) Joe would have done something that would have activated the device, precluding him from being morally responsible for the outcome. Pereboom responds by offering the following account of robustness:

For an alternative possibility to be relevant per se to explaining why an agent is morally responsible for an action it must satisfy the following characterization: she could have willed something different from what she actually willed such that she has some degree of cognitive sensitivity to the fact that by willing it she thereby would be, or at least would likely to be, precluded from the responsibility she actually has. (Pereboom, 2009, p. 110)

Let us call the requirement that the agent has some degree of cognitive sensitivity the 'cognitive sensitivity condition.' To motivate this condition on robustness, consider a situation offered by Pereboom (2012, p. 299). Suppose that the only way Joe could have

avoided deciding to evade his taxes is by taking a sip of coffee from his cup. Unbeknownst to him, someone has laced his coffee with a drug that induces compliance with the tax code. As things actually happen, Joe never takes a sip of his coffee and he decides to evade his taxes. Although Joe has an alternative possibility available to him—he could take a sip of his coffee—and performing this alternative act would preclude him from moral responsibility, this alternative is irrelevant to whether he is morally responsible. Joe has no cognitive sensitivity at all to the fact that taking a sip of his coffee would allow him to avoid moral responsibility. Joe's situation is similar to that of Tal's in McKenna's Needed Medication case. Although it is within Joe's voluntary control to avoid cheating on his taxes, there is simply no reason available to Joe for him to consider this option.

Given the cognitive sensitivity condition on robustness, Pereboom thinks it is clear that the possibility that Joe enters the buffer zone by vividly imagining his punishment is not robust. Joe has no idea that Black has implanted a device into his brain, much less that it would be activated by his voluntarily imagining the punishment for tax evasion. He has no cognitive sensitivity whatsoever to the fact that by imagining his punishment he would avoid responsibility for his decision. Moreover, Joe has no reason to believe that imagining his punishment would make it likely that he would not decide to evade his taxes. After all, Pereboom says, “one might imagine that he believes that achieving this level of attentiveness is compatible with his freely deciding to evade taxes anyway, and that he has no reason to suspect otherwise” (Pereboom, 2009, p. 114). For these reasons, Pereboom concludes that Joe is morally responsible for his decision to evade taxes but lacks robust alternatives to that decision.

### 3.8. Criticism of Buffer Cases

However, as presented, Tax Evasion (3) is problematic. It includes the following two incompatible claims:

(A) It is causally necessary for Joe's failing to decide to evade taxes that he imagine being fairly severely punished for evading taxes with a certain degree of vividness.

(B) Joe's failure to vividly imagine his punishment to the requisite degree at any given moment is not a causally sufficient condition for Joe's deciding to evade taxes.

Given (B), it cannot be the case Joe's vividly imagining his being severely punished to the requisite degree is necessary for his failing to decide to evade taxes. After all, (B) allows that at any given moment Joe might fail to enter the buffer zone and also fail to decide to evade taxes. At each moment before Joe makes a decision he neither decides to evade taxes nor vividly imagines the punishment he would face if caught. Indeed, there is nothing that prevents Joe from failing to make a decision into the indefinite future. Thus, contrary to (A), Joe could fail to decide to evade taxes without imagining his punishment if he is caught.

To avoid this difficulty, we must introduce a deadline:

(A') It is causally necessary for Joe's failing to decide to evade taxes *by a particular time t* that he imagine being fairly severely punished for evading taxes with a certain degree of vividness.

(B') Joe's failure to vividly imagine his punishment to the requisite degree at any given moment *prior to t* is not a causally sufficient condition for Joe's deciding to evade taxes.

I think this captures Pereboom's intention with the Tax Evasion cases. The idea is that Joe is morally responsible for deciding to evade taxes by the deadline *t*, and Joe could not have failed to decide to evade taxes by *t*.

However, this alteration introduces a new difficulty. How is the scenario to be designed so that Joe is both guaranteed to decide to evade taxes by the deadline  $t$  and is morally responsible for his decision? Pereboom has offered a new case in which he introduces a time limit such that the agent will have made a decision by a particular time.<sup>17</sup> He calls this new case Tax Cut:

Jones can vote for or against a modest tax cut for those in his high-income group by pushing either the ‘yes’ or the ‘no’ button in the voting booth. Once he has entered the voting booth, he has exactly two minutes to vote, and a downward-to-zero ticking timer is prominently displayed. If he does not vote, he will have to pay a fine, substantial enough so that in his situation he is committed with certainty to voting (either for or against), and this is underlain by the fact that the prospect of the fine, together with background conditions, causally determines him to vote. Jones has concluded that voting for the tax cut is barely on balance morally wrong, since he believes it would not stimulate the economy appreciably, while adding wealth to the already wealthy without helping the less well off, despite how it has been advertised. He is receptive and reactive to these general sorts of moral reasons: he would vote against a substantially larger tax cut for his income group on account of reasons of this sort, and has actually done so in the past. He spends some time in the voting booth rehearsing the relevant moral and self-interested reasons. But what would be required for him to decide to vote against the tax cut is for him to vividly imagine that his boss would find out, whereupon due to her political leanings she would punish him by not promoting him to a better position. In this situation it is causally necessary for his not deciding to vote for the tax cut, and to vote

---

<sup>17</sup> Pereboom provides this case in response to an objection raised by Ginet (2002), Franklin (2011b), and Palmer (2011) known as the “Timing Objection.”

against it instead, that he vividly imagine her finding out and not being promoted, which can occur to him involuntarily or else voluntarily by his libertarian free will. Jones is sensitive to the fact that imagining this punishment scenario will put him in a motivational position to vote against. But so imagining is not causally sufficient for his deciding to vote against the tax cut, for even then he could still, by his libertarian free will, either decide to vote for or against (without the intervener's device in place). However, a neuroscientist [Black] has, unbeknownst to him, implanted a device in his brain, which, were it to sense his vividly imagining the punishment scenario, would stimulate his brain so as to causally determine the decision to vote for the tax cut. Jones's imagination is not exercised in this way, and he decides to vote in favor while the device remains idle. (Pereboom, 2013, pp. 429-430)

The relevant difference between this case and Tax Evasion (3) is that in Tax Cut the circumstances causally determine that the agent will make a decision by a certain time. Because of the deadline and the penalty for failing to vote by the deadline, Jones is committed to making a decision to vote for or against the tax cut before time runs out. Since Jones will make a decision to vote either for or against the tax cut, and since he cannot decide to vote against the tax cut, Black is able to ensure that Jones will decide to vote 'yes' at some point within a two-minute voting window, but Black is not able to ensure that Jones decides to vote 'yes' at any specific moment within the time frame. Thus, although it is within Jones's power to avoid being morally responsible for deciding to vote 'yes' at any given moment, it is not within Jones's power to avoid deciding to vote 'yes' *by the deadline*. Pereboom concludes that Jones could be morally responsible for choosing to vote for the tax cut by the deadline even though he has no robust alternative possibilities.

Tax Cut avoids the objections raised against Brain Malfunction. Since Jones is causally determined to make a decision by the deadline, simply omitting to vote ‘yes’ by the deadline is not an alternative that is available to him. Neither is deliberating until time runs out. Unlike Brain Malfunction, the assumption of causal determinism is not problematic in Tax Cut. This is because, although Jones is causally determined to vote one way or another by the deadline, he is not causally determined to vote ‘yes’ by the deadline.

The chief difficulty with Pereboom’s approach, as I see it, is that it seems quite plausible that Jones’s moral blameworthiness for deciding to vote ‘yes’ by the deadline is derivative of his blameworthiness for failing to do his reasonable best to avoid making that decision. What Jones should have done, it might be argued, is imagine the punishment scenario with the requisite level of vividness. After all, Jones understands that voting in favor of the tax cut is morally wrong and that the first step he must take if he is to avoid deciding to vote in favor of the tax cut is to vividly imagine the punishment he would receive if his boss discovered that he voted ‘yes.’ Given these features of the case, and provided that his moral reasoning is unimpaired, Jones will recognize (or should recognize) that he is morally obligated to vividly imagine his punishment. Had he done that, the neuroscientist would have intervened and he would not be morally responsible for his decision to vote in favor of the tax cut. But Jones doesn’t take that first step towards doing what he recognizes is the right thing to do, and it is in virtue of this failure to act that Jones is morally blameworthy for his decision.<sup>18</sup>

I shall consider three objections to this response. First, the alternative possibility available to Jones—imagining his potential punishment with adequate vividness—fails to

---

<sup>18</sup> Widerker (2006) and Elzein (2013) respond this way to Pereboom’s Tax Evasion cases. But the point applies equally well to Tax Cut.

meet one of the conditions in Pereboom's account of robustness. Pereboom says that if an alternative possibility is robust, then the agent must be cognitively sensitive to the fact that willing the alternative would preclude the agent, or at least *make it likely* that the agent is precluded, from the responsibility she actually has. Although Jones recognizes that he must perform the intermediate action before he is able to fail to vote in favor of the tax cut, he might not be cognitively sensitive to the fact that performing such an action makes it likely that he will fail to vote 'yes.' Indeed, it may well be that were the counterfactual intervener not present, even if Jones were to imagine his punishment, he would still be unlikely to fail to vote in favor of the tax cut. And given that Jones is unaware of the presence of the counterfactual intervener, he may well believe that imagining his punishment would not make his failing to vote in favor likely. So, on Pereboom's view, although Jones does have an alternative that would preclude him from moral responsibility (given the presence of the counterfactual intervener), such an alternative is not robust.

Nadine Elzein (2013) offers a potent challenge to Pereboom's criterion for robustness. She provides the following case:

Suppose Suzy...knows that Billy is absolutely certain to die if she does not give him the injection, but that there is a 45% chance that he will recover fully if she does administer the injection. Suppose she could easily administer the injection, but she chooses not to, as she is reading a rather gripping airport novel and would rather not put it down. (Elzein, 2013, p. 560)

On Pereboom's account of robustness, Suzy's alternative is not robust since it is not *likely* to prevent Billy's death. But, intuitively, it is a morally significant alternative. What matters, it seems, is not that Suzy's performing the alternative action makes it *likely* that Billy will recover, but that doing so makes a recovery *more likely* than it would be given the course of



action she actually pursues. What this case shows, says Elzein, is that “alternatives seem important so long as they are *comparatively* likely to result in better outcomes. We are much less interested in the question of whether they are likely to result in better outcomes full stop” (Elzein, 2013, p. 560). And this is true even when the “better” outcome is one in which the agent is not subject to blame. Elzein offers the following case:

Suppose Suzy has contracted a virus, which she knows will make her feel more easily irritated than usual, and will reduce her level of concern for the moral demands of her situation (though she will still understand those demands well enough to count as a responsible agent). In fact, some very knowledgeable scientists have profiled her personality, and they calculate that it is almost 100 % certain that she will get irritated and murder Billy (who is in her care, and can be annoying) unless she takes a vaccine, which will reduce the chances of this wrongdoing to just 55 %. Suppose that she knows all of this, but decides not to take the vaccine, once again because this would interrupt her gripping airport novel. And finally, suppose she does in fact end up killing Billy. (Elzein, 2013, p. 560)

Although taking the vaccine was unlikely to preclude Suzy from blame, the availability of such an alternative seems relevant to our assessment of her moral blameworthiness because, by accessing this alternative, she would have been much more likely to avoid committing the morally wrong act for which she is blamed, and she was aware of this fact. Likewise, with regard to Tax Cut, as far as Jones understands the situation, choosing the alternative may not *make it likely* that he is precluded from blame, but doing so certainly makes it *more likely* that he will avoid being blameworthy, and Jones is well aware of this fact.

Additionally, Elzein points out the absurdity of requiring *absolute likelihood* for robustness instead of *comparative likelihood*. Suppose an alternative must be 75% likely to

preclude an agent from moral responsibility for it to count as a robust alternative. On Pereboom's account, an alternative that raises the likelihood of exempting an agent of moral responsibility from 70% to 75% would count as a robust alternative, but an alternative that raises the likelihood from 0% to 70% would not. But, of course, such a result seems quite strange. All of these considerations suggest that Pereboom's requirement that an alternative makes it likely (full stop) that an agent is precluded from blame for that alternative to count as robust is mistaken.

A second objection is that Tax Cut differs in significant ways from paradigmatic cases of derivative responsibility. Pereboom offers the following case of derivative responsibility (2014, p. 18). Biff understands that if he is intoxicated, he will not be able to avoid being abusive to his friends. He decides to get drunk anyway and while drunk assaults one of his friends, Marty. Although Biff does not satisfy the general conditions on moral responsibility while he is drunk, says Pereboom, he does satisfy them before he decides to get drunk. Thus, if Biff is morally responsible for assailing Marty, his responsibility is derivative of his being non-derivatively responsible for his earlier decision to get drunk.

The crucial difference between Biff and Jones, says Pereboom, is that "Biff has knowingly placed himself in a position in which some of the uncontroversial general conditions on non-derivative moral responsibility will fail to be satisfied at relevant subsequent times" and Jones has not (2014, p. 19). Once Biff is drunk, he is no longer able to respond adequately to the moral demands of his situation. Jones, however, could at any moment (until the deadline) vividly imagine his punishment scenario, enabling him (he thinks) to respond to the moral demands of the situation as he ought to. Thus, although Biff and Jones each do something morally blameworthy, and each could have done something earlier that would have enabled them to avoid performing the morally blameworthy act,

Jones continually has the opportunity to do that thing up until the moment he performs the morally blameworthy act and Biff does not.

Although Pereboom has drawn out a relevant difference between *one* paradigmatic case of derivative responsibility and Tax cut, there are other cases of derivative responsibility in which this difference cannot be so easily drawn. Consider the following. Suppose Mr. Toad drives his automobile over the narrow lanes of the countryside as fast as he can. He knows that at such speeds, he would not be able to avoid anyone should he or she appear on the road in front of him. Mr. Toad also knows that he could reduce his speed at any moment before anyone appears on the road, and that this would give him enough control over the car to avoid him or her if he so chose. Moreover, Mr. Toad knows that he is driving at the same time that Mr. Badger walks along the lane. As things happen, Mr. Toad, delighted to be moving so quickly over the lane, does not reduce his speed, Mr. Badger appears on the road in front of him, and Mr. Toad hits and injures Mr. Badger. Mr. Toad's hitting Mr. Badger with his car, it seems to me, is a clear-cut case of derivative responsibility, yet Mr. Toad continuously has an action available to him (up until the moment Mr. Badger appears) that would enable him to respond adequately to the moral demands of his situation (i.e., do not hit and injure Mr. Badger): he could reduce his speed. Mr. Toad, like Jones, is morally responsible on the basis of his continued failure to do what he knows he needs to do to gain access to the ability to avoid moral wrongdoing.<sup>19</sup>

---

<sup>19</sup> Elzein (2013) makes the same point by suggesting that we modify a case like Biff's such that the agent has the same ability that Jones has. Suppose Biff is under the influence of a new drug that puts him in a state of euphoria without impairing his judgment, but makes him abusive to his friends in the same way that getting drunk does. Biff knows that if he remains under the influence of this drug, he will assault Marty. He also knows that if he were to take a sip from the cup of coffee in his hand he would immediately no longer be under the effects of the drug. As things happen, Biff does not take a sip of his coffee and he assaults Marty.

This response helps us deal with a third objection. Pereboom argues that the derivative responsibility response to Tax Cut is dialectically unsatisfying since it appeals to a leeway position for support. As he sees it, the reason for thinking that Jones is only non-derivatively morally responsible for his failure to vividly imagine his punishment is because it is only relative to this decision that Jones has a robust alternative. Thus, the alternative-possibilities schema would go like this: “the agent is non-derivatively morally responsible for acting or refraining at a particular time only if a robust alternative possibility relative to so acting or refraining is available to her at that time, and all other moral responsibility is derivative of such non-derivative responsibility” (Pereboom, 2014, p. 19). However, this schema begs the question by presupposing that robust alternatives are necessary for non-derivative moral responsibility.

But consideration of Mr. Toad’s case allows us to claim that Jones is derivatively morally responsible for voting in favor of the tax cut without appealing to the above schema for support. If one grants that the case of Mr. Toad is a paradigmatic case of derivative culpability, and that the situation is relevantly similar to the situation Jones finds himself in, then one has good reason to think that Jones’s moral blameworthiness for voting in favor of the tax cut is also derivative. There is no need to appeal to the presence of robust alternatives in support of this conclusion. One need only look at this paradigmatic case of derivative moral responsibility and note its similarity with Jones’s position in Tax Cut.

Indeed, with only minimal adjustments, the case of Mr. Toad can be turned into a buffer style Frankfurt example similar to Tax Cut by inserting a counterfactual intervener. Suppose the local weasels have installed a device in Mr. Toad’s automobile that would activate and increase the automobile’s speed if Mr. Toad were to begin to reduce his speed at any time before Mr. Badger appears in the lane. However, Mr. Toad never begins to

reduce his speed, the device never activates, and Mr. Toad hits and injures Mr. Badger with his automobile. Despite the addition of the counterfactual intervener, the case remains a clear-cut example of derivative responsibility. Thus, there seems to be nothing about Tax Cut that makes it relevantly different from the case of Mr. Toad such that one would say Jones is non-derivatively culpable while Mr. Toad is derivatively culpable.<sup>20</sup>

To clarify, the response to Pereboom's third objection goes like this. The Mr. Toad scenario is a clear-cut case of derivative responsibility even though Mr. Toad has an action available to him throughout the scenario that would allow him to respond appropriately to the moral demands of his situation. Introducing a counterfactual intervener doesn't seem to change this. Thus, one may argue that the agent in Pereboom's buffer case is derivatively morally responsible because the case is relevantly similar to a clear-cut case of derivative responsibility—the Mr. Toad case—without having to appeal to the alternative-possibilities schema that Pereboom argues is question begging.

### ***3.9. Concluding Remarks***

The prospects for finding a modal counterexample to PAP(CC) are not good. As we have seen, the most promising attempts are deeply flawed. I think this partly explains why it remains controversial whether Frankfurt examples succeed, even after more than forty years have passed since the publication of Frankfurt's article, "Alternate Possibilities and Moral Responsibility." The failure to provide a clear cut modal counterexample to PAP(CC) after all that time has led some to conclude that Frankfurt's argument fails. However, as we saw in Chapter 2, Frankfurt examples don't need to succeed in a modal strategy against PAP(CC) in order to succeed in demonstrating that PAP(CC) is false. Some can be used in

---

<sup>20</sup> See Elzein (2013) and Lockie (2014) for various other responses to Pereboom's objection.

a non-modal strategy. However, even though PAP(CC) is false, it does not follow that PAP is also false. A number of independent arguments in support of PAP have been offered. I shall consider two of them in the next chapter.

## Chapter 4. Two Arguments for the Principle of Alternative Possibilities

In Chapter 2 I argued that PAP(CC) is false. However, since PAP does not entail PAP(CC), it remains possible that PAP is true even if PAP(CC) is false. In this chapter I shall consider two arguments for PAP. The first turns on a principle that David Widerker calls the “Principle of Alternative Expectations.” The second involves Kant’s maxim that “ought” implies “can.” I shall argue that both of these arguments for PAP fail.

### *4.1. Widerker’s Defense of PAP*

Widerker (2000, 2003, 2005) offers an interesting defense of the Principle of Alternative Possibilities:

PAP: Necessarily, an agent is non-derivatively morally responsible for performing an action only if he could have avoided performing it.<sup>1</sup>

As Widerker sees it, Frankfurt’s argument against PAP depends on the following two claims:

IRR: There may be circumstances in which an agent performs some action such that, although the circumstances make it impossible for the agent to avoid performing the action, they in no way bring it about that the agent performs the action.

M: It is possible that an agent in an IRR-situation (i.e., a scenario of the sort described in IRR) who performed a morally wrong action is non-derivatively morally blameworthy for performing that action.

If these two claims are true, then PAP is false. It is possible for an agent to be morally responsible even though he couldn’t have done otherwise. This is where, Widerker says, the

---

<sup>1</sup> Recall from Chapter 1 that moral responsibility in PAP is limited to the sort of responsibility required for moral blameworthiness.

Frankfurt examples enter the discussion. They are intended to be metaphysically possible scenarios that demonstrate the truth of IRR. Recall the prior sign Frankfurt example from Chapter 2, Bribe:

Smith is pushing a bill through Congress that puts severe restrictions on neuroscientific research. Congressman Jones has been offered a large sum of money by the League of Nefarious Neuroscientists (LNN) to vote against the bill. Black, the founder of the LNN, would prefer that Jones accept the bribe on his own. However, not wanting to take any chances, Black has, unbeknownst to Jones, implanted a special device into Jones's brain. Lucky for Black, Jones has a tell that reliably indicates what he will do. If Jones is going to decide to accept the bribe on his own by a certain time,  $t_3$ , he will exhibit an involuntary facial twitch,  $F$ , prior to  $t_1$ . If Jones exhibits  $F$  prior to  $t_1$ , Black will not activate the device. However, if Jones is not going to decide to accept the bribe on his own by  $t_3$ , he will not display  $F$  prior to  $t_1$ , and Black will activate the device at  $t_2$ , causing Jones to choose to accept the bribe by  $t_3$ . As things actually happen, Jones exhibits  $F$  before  $t_1$ , Black never activates the neural device, and Jones decides to accept the bribe by  $t_3$ .

The inclusion of Black (and his device) is intended ensure that Jones could not have avoided deciding to accept the bribe. But, as things actually happen, Black never intervenes and Jones decides to accept the bribe on his own. Thus, it appears that the circumstances ensure that Jones performs a particular action without in any way bringing it about that he does perform that action.

Much of the debate over Frankfurt examples has centered on whether it is possible to provide a coherent IRR-situation. However, rather than pursue that strategy in defense of PAP, Widerker targets M. He grants for the sake of argument that it is possible to produce



an unproblematic example of an IRR-situation. To keep things simple, let us assume that Bribe is such an example.<sup>2</sup> Jones did what he did for reasons of his own without being coerced or causally determined to act. Widerker argues that no agent in an IRR-situation could ever be morally responsible for what he has done.

Widerker has construed Frankfurt's argument as a modal argument against PAP.<sup>3</sup> However, in Chapter 2 I argued that the best strategy against PAP is to argue that it is undermotivated by targeting PAP(CC), a principle that grounds the plausibility of PAP.

PAP(CC): When an agent is non-derivatively morally responsible for an action, he is so partly in virtue of having been able to have done otherwise.

Without this principle or further support for PAP, there is little reason to believe PAP is true. Recall the non-modal argument against PAP(CC):

The complete blockage version of Bribe (from Chapter 2) shows that it is possible that the fact that an agent is unable to do otherwise is irrelevant to an explanation of what the agent did and why he did it. If a fact is irrelevant to an explanation of what an agent did or why an agent did it, then that fact has no bearing on the agent's moral responsibility. Thus, it is possible that the fact that an agent is unable to do otherwise has no bearing on the agent's moral responsibility. Therefore, PAP(CC) is false; the availability of alternative possibilities is not part of the conceptual fabric of moral responsibility.

Notice that the non-modal argument against PAP(CC) (and the resulting indirect argument against PAP) does not require M. The Frankfurt defender need not claim that the agent in an IRR-situation *is* morally responsible. Thus, one might think that even if

---

<sup>2</sup> Of course, it is clear from the discussion in Chapter 2 that Bribe is *not* an unproblematic instance of an IRR-situation.

<sup>3</sup> Widerker has also formulated a version of his argument that can be used to target PAP(CC). See Widerker (2005).

Widerker is successful in showing that M is false, this will not matter for the Frankfurt defender who adopts the strategy outlined above. However, Widerker's attack on M is relevant in two ways. First, in Widerker's criticism of an argument offered by Frankfurt in support of M, Widerker attacks a key principle that I use in the non-modal argument against PAP(CC). Thus, if his rejection of this principle is successful, the non-modal argument outlined above fails. Second, Widerker offers an argument against M, the W-defense, that if successful, can be used to support PAP directly.

#### ***4.2. Targeting the Irrelevance Principle***

Widerker draws an argument for M from the following passage of Frankfurt's original article on PAP:

The fact that a person could not have avoided doing something is a sufficient condition for his having done it. But as some of my examples show, this fact may play no role whatever in the explanation of why he did it. It may not figure at all among the circumstances that actually brought it about that he did what he did...The circumstances that made it impossible for him to do otherwise could have been subtracted from the situation without affecting what happened or why it happened in any way.

When a fact is in this way irrelevant to the problem of accounting for a person's action it seems gratuitous to assign it any weight in the assessment of his moral responsibility. Why should the fact be considered in reaching a moral judgment concerning the person when it does not help in any way to understand either what made him act as he did or what, in other circumstances, he might have done? (Frankfurt, 1969, pp. 836-837)

Widerker says that Frankfurt's argument turns on the following principle (let us call it the "Irrelevance Principle"):

[IP:] If a fact is irrelevant to the explanation of why the agent performed a certain act, then such a fact has no bearing on the agent's moral responsibility for the act. (Widerker, 2003, p. 60)

Since an agent's inability to do otherwise is irrelevant to an explanation of why the agent performed the action she did in an IRR-situation, it follows from IP that her inability to do otherwise has no bearing on her moral responsibility for her action. Therefore, "in the absence of any other excusing factors, M is the case" (Widerker, 2003, p. 60).

IP is a key premise in my non-modal argument against PAP(CC).<sup>4</sup> So, if the principle turns out to be false, that is a problem for my argument.<sup>5</sup>

Widerker offers three counterexamples to IP. In the first, a person, Smith, mistakenly believes that he can prevent a hurricane from destroying a city by reciting an incantation. Smith decides not to recite the incantation, and the hurricane destroys the city. The fact that reciting the incantation will not prevent the city from being destroyed plays no part in why Smith decides not to recite the incantation. However, this fact *does* seem relevant to why Smith is not morally blameworthy for failing to prevent the disaster. The reason why Smith

---

<sup>4</sup> My argument actually turns on a slightly different principle, IP\*, which I shall explain momentarily.

<sup>5</sup> There is another argument that could be offered using the non-modal strategy, one that turns on a principle suggested in Frankfurt's (1969) original article:

An agent can cite a fact as a basis for exoneration only if that fact reveals either that she does not act from a culpable motive, or that her action issued in a causally deviant fashion. (McKenna, 2008, pp. 773, footnote omitted)

Widerker has challenged this principle as well (2003, p. 62); McKenna has responded (2008, pp. 780-783).

is not blameworthy is, at least in part, that reciting the incantation will have no effect on whether the hurricane destroys the city.<sup>6</sup>

In Widerker's second example, a person, Green, does not want to go to work and discovers that he is sick. He decides not to go to work, but the fact that he is sick plays no role in why he decides not to go to work. Instead, he decides not to go because he does not feel like it. For this to work as a counterexample to IP, as Widerker intends, it must be the case that Green is not blameworthy for deciding not to go to work, and the reason he is not blameworthy is that he is sick. In this way then, the fact that Green is sick is irrelevant to why he decides not to go to work, but it *is* relevant to why we do not hold him blameworthy.

In a third case, one Widerker believes to be the strongest, Green harms another person for some selfish reason and knows that in doing so he is acting immorally, but he does not harm the person *because* he knows that doing so is morally wrong. Widerker says that the fact that Green knows what he is doing is wrong has some bearing on his being blameworthy for the act even though the fact plays no role in an explanation of why he performed the act.

Frankfurt (2003) has responded to the first and third examples. Regarding the first case, Frankfurt agrees with Widerker that Smith is not blameworthy for failing to prevent the disaster. However, he says that this has nothing to do with PAP. Smith is not blameworthy for not preventing the disaster because he performed no such act. "Nothing that Smith did can be correctly described as an act of not preventing, or of refraining from preventing, or of failing to prevent the disaster. It is this, and not his inability to prevent the disaster, that

---

<sup>6</sup> Widerker grants that Smith may be held blameworthy for *something* in this situation—perhaps for not trying to prevent it, or for not deciding not to prevent it—but he says that Smith is not blameworthy for failing to prevent it.

decisively precludes his being held morally responsible for not preventing, or for refraining from preventing, or for failing to prevent it” (Frankfurt, 2003, p. 341).

As for the third case, Frankfurt grants that the fact that Green knew that he was acting immorally has a bearing on his blameworthiness, and Frankfurt grants that Green’s awareness of this fact played no role in motivating him to perform the action he did; however, the fact that Green believed the act to be immoral *is* relevant to the explanation of what Green did. Frankfurt says,

A full explanation of what Green did must provide more than just a statement that his motives were selfish. It must also report that he acted as he did because he cared more about attaining his selfish goals than he cared about avoiding immorality. This is relevant to his blameworthiness because it bears on what sort of action it was that he took himself to be performing. If he had performed the same act while believing it to be morally neutral, we would judge his conduct differently. (Frankfurt, 2003, p. 343)

Frankfurt makes a good point. However, as far as I can tell, he has not adequately responded to Widerker’s objection. Widerker’s case attempts to show that features of a situation that are relevant to whether an agent is morally blameworthy might play no part in explaining why the agent performed the action he did. Frankfurt grants that Widerker’s third example shows this. And he is right to do so. However, if we grant this, then IP is false.

Consider another example. Recall Locke’s locked room case from Chapter 2. A man is inside a room with locked doors, but he does not know that the doors are locked. His child is crying for help outside the door. He decides to stay in the room, knowing that in doing so he is acting immorally. If IP is true, then from the fact that the child crying outside the door is explanatorily irrelevant to why the man decided to stay inside the room it follows that the

crying child is irrelevant to why the man is morally responsible for his decision.<sup>7</sup> But of course, this is false. The crying child *is* relevant to an explanation of why the man is morally responsible.

This example and Widerker's third example show that IP is false. However, as I see it, Widerker has misidentified the principle that Frankfurt's argument turns on. The key principle is the following:

IP\*: Necessarily, if a fact is irrelevant to a full explanation of *what* action an agent performed and *why* he performed it, then such a fact has no bearing on the agent's moral responsibility for the act.

In an IRR-situation, the circumstances that ensure that an agent performs a particular action do not themselves bring about that action. They are not relevant to a full explanation of *what* the agent does or *why* the agent does it. This is why, in his original article, Frankfurt says that these circumstances "could have been subtracted from the situation without affecting *what happened* or *why it happened* in any way" (Frankfurt, 1969, my emphasis). This, I think, is why Frankfurt responds to Widerker's third example as he does, by talking about what is relevant to a full explanation of *what* the agent did. Frankfurt had in mind IP\*, not IP.

What about Widerker's second example? It is difficult to see why Frankfurt should grant that Green is not blameworthy for deciding not to go to work. The reason Green decides not to go to work is that he doesn't feel like it, and this fact seems to be relevant to whether he is blameworthy. The fact that Green is sick seems irrelevant not only to why Green decides as he does but also to whether he is morally responsible for his decision.

---

<sup>7</sup> I wish to thank Tomás Bogardus for offering this example during my presentation at the SoCal Philosophy Conference, University of California, San Diego, October 18, 2014.

After all, were the situation altered so that Green was not sick, he would still decide as he did for the same reasons that he did. Of course, it might be the case that no one but Green knows the true reason why he stayed home from work, and therefore, no one may actually be in a position to blame Green. But it does not follow from this that Green is not worthy of blame.<sup>8</sup>

The key principle my non-modal argument against PAP(CC) turns on is IP\*, not IP. IP\* is intuitively appealing. None of the cases offered by Widerker give us reason to doubt its truth. Moreover, IP\* is neutral with regard to the debate between compatibilists and incompatibilists. Someone might grant IP\* yet maintain that moral responsibility is incompatible with causal determinism. For example, those committed to an actual-sequence approach to moral responsibility might do so. On the actual-sequence view, what matters for assessing moral responsibility is found in the actual sequence of events: the action that was performed and how it was performed. Whether moral responsibility is consistent, on this view, with causal determinism is another matter. Source incompatibilists maintain that in order for an agent to be morally responsible for her action she must be the ultimate source of that action, and they hold that if causal determinism is true then no one is the ultimate source of their actions.<sup>9</sup> Thus, a source incompatibilist might grant that alternative possibilities are not required for moral responsibility, yet maintain that causal determinism rules out an agent's moral responsibility.

---

<sup>8</sup> McKenna makes this point in his reply to Widerker (2008, pp. 779-780).

<sup>9</sup> See, for instance, Hunt (2000), Pereboom (1995, 2001, 2014), Stump (1996), and Zagzebski (2000).

### 4.3. The W-defense

With Widerker's objections to the Irrelevance Principle out of the way, I now turn to his argument for PAP. Recall that he assumes, for the sake of discussion, that the Frankfurt example under consideration (in this case, Bribe) is an IRR-situation; Jones acted for reasons of his own and the circumstances entailing he couldn't do otherwise did not in any way bring it about that he performed the action that he did. He then presents his challenge to Frankfurt:

[S]ince you, Frankfurt, wish to hold him blameworthy for his decision to [accept the bribe], tell me *what, in your opinion, should he have done instead?* Now, you cannot claim that he should not have decided to [accept the bribe], since this was something that was not in Jones's power to do. Hence, I do not see how you can hold Jones blameworthy for his decision to [accept the bribe]. (Widerker, 2003, p. 63)

When Widerker asks what Jones *should* have done instead, he is not asking what Jones *ought* to have done in the sense of asking what Jones was *obligated* to have done. He grants that Jones was *obligated* not to decide to accept the bribe; Jones *ought* not have decided to accept the bribe.<sup>10</sup> Rather, he is asking what it would have been *reasonable to expect* Jones

---

<sup>10</sup> See Widerker (2005, pp. 303-304). He contends that his argument does not require the maxim "ought" implies "can." In reply to Fischer's claim that "rejecting the Maxim completely disarms the W-defense" (2006a, p. 210), he says,

There are well-known counterexamples to this principle. Consider the following one. Suppose that Jones borrows from Smith a rare copy of *Principia Mathematica* and promises to return it by October 10, 2003. On October 9, on his way to return the book to Smith, Jones gets robbed (due to no fault of his own), is unable to keep the promise, and does not keep it. Is he blameworthy for not having kept the promise? Obviously not! Has he failed to fulfill his obligation to give the book back to Smith by the said date? Certainly! He promised to return it by a certain date, and did not. Hence, "ought" does not (always) imply "can." Note that both these verdicts are compatible with PAE, since in the circumstances, it would be unreasonable to expect Jones to return the book by October 10. (Widerker, 2005, pp. 303-304)

If Widerker's argument *does* require the maxim, despite his claim to the contrary, then his argument stands or falls with the maxim. Later in this chapter I shall consider an argument



to have done under the circumstances. More generally, Widerker is asking what one could *reasonably expect* an agent in an IRR-situation to have done given the circumstances. Call this the “W-question.” Widerker thinks that if the answer to the W-question is “nothing,” then it is not reasonable to hold the agent blameworthy for what he did. He explains:

When we consider someone morally blameworthy for a certain act, we do so because we believe that morally speaking he should *not* have done what he did. This belief is essential to our moral disapproval of his behavior. Sometimes, however, such a belief may be unreasonable, for example, in a situation in which it is clear to us that the agent could not have avoided acting as he did. To expect in that situation that the agent should not have done what he did is to expect him to have done the impossible. By implication, considering him blameworthy because he has not fulfilled this unreasonable expectation would be unreasonable. (Widerker, 2003, pp. 63, footnotes omitted)

As Widerker sees it, to reasonably hold Jones blameworthy for deciding to accept the bribe we must believe that he should not have accepted the bribe. But it is unreasonable to believe that Jones should not have accepted the bribe if he could not have done otherwise. For, that would be to expect him to have done the impossible. Therefore, we cannot reasonably hold Jones blameworthy for deciding to accept the bribe; M is false.

Widerker calls this argument the “W-defense” (which is short for “what-should-he-have-done-defense”). Widerker says that the W-defense suggests the following constraint on assigning moral blame, the Principle of Alternative Expectations:

---

for PAP that does depend on this maxim and I shall argue that there is good reason to reject the maxim.

PAE: Necessarily, if an agent S is non-derivatively morally blameworthy for performing an action A, then it would be morally reasonable to expect S not to A under the circumstances.<sup>11</sup>

With PAE in hand, Widerker formulates an argument for PAP:

(W1) Necessarily, if an agent S is non-derivatively morally blameworthy for performing an action A, then it would be morally reasonable to expect S not to A under the circumstances (PAE).

(W2) Necessarily, if S could not have avoided performing A, then it would be morally unreasonable to expect him not to A (on pain of expecting him to have done the impossible).

(W3) Therefore, necessarily, if S could not have avoided performing A, then S is not non-derivatively morally blameworthy for performing A.

The conclusion, (W3), is logically equivalent to PAP.<sup>12</sup> In the next section I shall consider several responses to the W-defense.

#### ***4.4. Replies to the W-defense***

Eleonore Stump (2003) has challenged the W-defense by arguing that there is a good answer to the W-question. She says that there is something else we could have reasonably expected Jones to have done. Although it was not up to Jones *whether* he decided to accept the bribe, *how* he decided to accept the bribe was up to him. His choice to accept the bribe

---

<sup>11</sup> By ‘morally reasonable,’ Widerker means “reasonable for someone who is morally competent and knows all the facts that are relevant for determining whether (under the circumstances) it would be reasonable to expect the agent not to have done A” (Widerker, 2003, pp. 72 , endnote 27).

<sup>12</sup> This is so because the sort of responsibility we are interested in here is the sort of responsibility required for moral blameworthiness.

was either coerced or it wasn't, and it was up to him which it was. Therefore, Stump says, the answer to the W-question is that "Jones should have brought it about that his morally culpable act is one which the intervener and not Jones himself is culpable for causing" (2003, pp. 151, footnote omitted).

An initial worry for Stump's response to the W-question is that it appears to introduce an alternative possibility into the Frankfurt example. If Stump's response is correct, then an alternative possibility seems to be available to Jones: he could have brought it about that his decision to accept the bribe was coerced. Stump anticipates this reply and provides a response. She says that although there are alternative *modes* available to Jones in which his decision to accept the bribe comes about, there are no alternative *actions* available to Jones. And these alternative *modes* that are available to Jones do not support PAP since PAP is a principle about alternative *actions*.<sup>13</sup>

How could Jones be morally responsible if he couldn't have done otherwise? Stump says that what matters is that Jones originated his action.

On the response I have given to Widerker's question, it is true that it is up to Jones whether or not he does what is blameworthy. But it need not be up to Jones in virtue of

---

<sup>13</sup> One might respond to Stump by claiming that an agent has access to these alternative *modes* only if he has access to alternative actions. That is, the way that Jones is able to choose *how* he decides to shoot Smith is through performing some prior action. (This would be a mental action: deciding or resolving not to take the bribe.) So, rather than successfully responding to the W-defense, Stump's reply has just pushed the question of whether Jones could have done otherwise (or decided otherwise) back a step. Stump responds to this worry by pointing out that the way in which an action originates cannot always itself be another action, on pain of an infinite regress.

But the manner in which an action *A1* originates is not itself another action *A2* on the agent's part. If it were, presumably *A2* would also have a manner in which it originates, and so there would be an act *A3* which would be the agent's originating *A2*; and so on, *ad infinitum*. If, however, the manner in which an act originates is not itself an act and nonetheless is sufficient for moral responsibility (or lack of it), then an agent can have control over a sufficient condition for his being responsible without having control over which acts he performs. (Stump, 2003, p. 152)

the fact that alternative possibilities for action are open to him. If an act's originating indeterministically in an agent's own intellect and will is sufficient for moral responsibility, then moral responsibility can be tied just to the manner in which the act originates. (Stump, 2003, pp. 151-152)

Stump agrees with Widerker that there is no action one could reasonably expect Jones to have performed instead. But she doesn't think that there needs to be one for Jones to be reasonably held morally responsible. According to Stump, Jones is morally responsible (in part) because he originated his action. Had the act not originated with Jones—i.e., in his “intellect and will”—then Black would have intervened and Jones would not be morally responsible for his decision. Thus, Jones has control over the way he decides even if he couldn't have decided differently.

Stump's response will not appeal to those who are not already committed to the actual-sequence approach to moral responsibility. If we grant that PAP only concerns alternative *actions*, then the W-question becomes: what *action* could one reasonably expect Jones to perform instead of deciding to accept the bribe? Stump's answer to this question is the same as Widerker's: “nothing.” This response will not be attractive to those people, like Widerker, who do not accept the actual-sequence theory. However, as I shall argue momentarily, Stump's response is essentially correct.

Michael McKenna (2005, 2008) says that there is no reasonable answer to the W-question. He concedes that the W-defense cannot be directly refuted. “[W]hen we blame a person for a moral wrong, a clear implication is that our moral charge includes the demand that the person not have done that, that the person have acted as morality requires” (McKenna, 2008, pp. 783, footnote omitted). And this demand that the agent did something

other than what she did indicates that there is “*some* presumption that she could have done that thing” (2008, pp. 784, footnote omitted). McKenna says,

Widerker has tapped into an important moral intuition favoring a conception of free agency and moral responsibility in terms of alternative possibilities. Some arguing for Frankfurt’s thesis are reluctant to admit that there could remain residual intuitions of our moral thought that cannot be fully accommodated. But Widerker’s W-Defense has simply hit one of those intuitions spot on. It is only philosophically honest to acknowledge that...My assessment of the W-Defense is that it is such an impressive argument because it hits upon a raw nerve for the Frankfurt Defender. By way of the W-Defense, certain features of ordinary thought are brought into relief; Frankfurt’s conclusions *do* lead to an “irresolvable tension” in our thinking about blame and its implications (regarding what an agent should have done instead). (McKenna, 2008, pp. 784-785)

As McKenna sees things, the W-defense draws our attention to an aspect of our thinking about moral responsibility that cannot be accommodated if the only sort of freedom left to agents is that which remains in an IRR-situation.

However, despite this concession, McKenna argues that the Frankfurt defender can resist Widerker. Instead of focusing on what the agent in an IRR-situation *should* have done, McKenna says that we should look at what the agent *did* do. In responding to the W-question, “What should the agent have done instead?” McKenna answers, “Look at what the agent has done.” He calls this the “L-reply.” McKenna points out that the intuitive force behind the W-defense is counterbalanced by the intuitive force behind the L-reply. Those who claim that an agent in an IRR-situation could not be morally responsible must deny the intuition that “A person’s moral responsibility concerns what she does do and her basis for

doing it, not what else she could have done” (McKenna, 2008, pp. 785, footnote omitted). They must give up the intuition that “What is relevant to moral responsibility is the moral quality of the motive with which an agent acts, combined with the manner in which she brings her actions about” (McKenna, 2008, pp. 786-787, italics removed).

According to McKenna, both the W-defense and the L-reply reveal competing intuitions about moral responsibility. They leave us with an “irresolvable tension” in our moral thinking. Thus, neither the W-defense nor the L-reply is sufficient for accepting or rejecting the possibility of an agent’s being morally blameworthy in an IRR-situation. Each side of the debate can point to an intuition that the other side is unable to accommodate. So, where do we go from here? As McKenna sees it, each side must appeal to further considerations in favor of their respective positions.

Justin Capes (2010) offers another response to the W-defense. He thinks that McKenna gives away too much ground. Frankfurt defenders need not concede that their position leaves them with an “irresolvable tension” because they need not grant PAE. But before fleshing out this response Capes raises an initial problem for Widerker’s argument. The W-defense does not directly support PAE or PAP. The conclusion of the W-defense is that it would not be *reasonable to hold* an agent in an IRR-situation morally blameworthy. However, the antecedent of PAE is not that it is *reasonable to hold* the agent blameworthy, but rather, that the agent *is* blameworthy. As Capes points out, the fact that it would be unreasonable to hold a person morally blameworthy for an action does not imply that that person is not morally blameworthy. After all, the available evidence might be misleading. For instance, suppose Black wants to slap Smith at a party, but he doesn’t want to be blamed for slapping Smith. Before the party Black pays Jones to hold onto his arm and make it look like Jones is forcing Black’s hand to slap Smith. During the party Black slaps Smith while

Jones holds onto his arm to make it appear as though he is controlling Black's hand. In this situation, given the evidence available, it would be unreasonable for Smith or any other partygoer (except Jones and Black, obviously) to blame Black for slapping Smith.

However, Black is clearly morally blameworthy for slapping Smith.

If the W-defense is successful, Capes says, the constraint on ascriptions of moral blame it suggests is not PAE, but rather,

PAE\*: Necessarily, if an agent S is *reasonably held to be* non-derivatively morally blameworthy for performing A, then it would be morally reasonable to expect S not to have A-ed under the circumstances.

However, Capes objects that PAE\* (along with premise (W2)) does not entail PAP since PAP is a principle about whether an agent *is* morally blameworthy, not about whether an agent is *reasonably held* morally blameworthy.

Capes sets the matter aside. But I think Widerker could easily sidestep this worry, and this turns out to have important consequences for Capes' response to the W-defense. The description of an IRR-situation is supposed to provide a God's-eye view of the situation. It is intended to provide all of the details of the situation that are relevant to determining whether the principal agent is morally responsible for performing the action he does. Thus, in assessing moral responsibility, the issue isn't whether someone *does* reasonably hold the agent morally responsible, but whether someone who knows the relevant facts of the case *could*.<sup>14</sup> Thus, we might revise PAE in the following way:

---

<sup>14</sup> This point is made by David Palmer (2013a).

PAE<sup>†</sup>: Necessarily, if an agent S is non-derivatively morally blameworthy for performing A, then it *could* be morally reasonable to expect S not to have performed A under the circumstances.<sup>15</sup>

Premise (W2) would also need to be revised:

(W2') Necessarily, if S could not have avoided performing A, then it *could not* be morally reasonable to expect S not to have performed A under the circumstances (on pain of expecting him to have done the impossible).

Together, PAE<sup>†</sup> and (W2') entail (W3).

Capes targets PAE\*. The trouble with the principle, he says, is that it is unclear why the Frankfurt defender should grant it. Since Jones could not have done otherwise than accept the bribe, it would be unreasonable to expect him not to have accepted the bribe. But this is troublesome for the Frankfurt defender only if holding Jones blameworthy commits the Frankfurt defender to the expectation that Jones not have done what he did. But Capes argues that the Frankfurt defender is not committed to this expectation. He says,

Widerker claims that an expectation that a person not act as she did is essential to our moral disapproval of her behavior. The Frankfurt-defender, I am suggesting, has been given no reason to accept this claim. Why must we expect a person to do otherwise in order to disapprove morally of the person's behavior? It seems entirely possible to disapprove of a person's behavior and indeed to hold that person blameworthy for what she has done without there being an expectation on our part that she not have behaved in

---

<sup>15</sup> The W-defense could be revised (without only minor alterations) accordingly: For Jones to be morally blameworthy for deciding to accept the bribe it must be the case that if someone were acquainted with the relevant facts of the case she could reasonably believe that Jones should not have accepted the bribe. But it is unreasonable to believe that Jones should not have accepted the bribe if he could not have done otherwise—that would be to expect him to do the impossible. Therefore, Jones is not morally blameworthy for deciding to accept the bribe.



that way. To be sure, we will no doubt wish that the person not have behaved as she did, and we may judge that the person ought not have behaved in that way, but as I have argued, this needn't involve an expectation or demand on our part that the person not behave as she did. (Capes, 2010, p. 75)

Capes seems to think that Widerker's acceptance of PAE\* commits him to the following:

E: An expectation that a person not act as she did is essential to our moral disapproval of her behavior.

Capes argues that E is false. He says it is possible that we hold someone morally blameworthy for an action without having any expectation that she not have acted in that way. On this basis, then, he rejects PAE\*.

David Palmer (2013a) has replied to this objection. He says that the rejection of E does not directly refute PAE\*. The fact that we can hold people morally blameworthy without having any expectations of them is consistent with PAE\*. This is because E is stronger than PAE\*. Palmer explains,

All [PAE\*] requires is that in order for a person to be held blameworthy, it must be morally reasonable *to* expect that he not have acted as he did. That is, it must be the case that a morally reasonable expectation *could* be had of him, not that one *must* be had of him. So, it does *not* follow from [PAE\*] that a morally reasonable expectation *must* be had of the person in order for him to be blameworthy, that he cannot be blameworthy *without* such an expectation being had of him. With this in mind, pointing out that a person can be held blameworthy without any expectation being had of him is perfectly consistent with [PAE\*], and so *not* a good criticism of [PAE\*] or of Widerker's W-defense argument for PAP. (Palmer, 2013a, p. 558)

Palmer is correct in pointing out that PAE\* does not entail E. Unlike E, PAE\* does not require that a morally reasonable expectation *actually* be had of the agent in order to hold him morally blameworthy, but rather, that one *could* be had of him.

The problem with Palmer's reply, however, is that it is difficult to see why someone who accepts PAE\* would reject E. I grant that there is logical space here to make this move. But what would motivate someone to make it, apart from a prior commitment to PAP? After all, making this move would commit one to saying that it is reasonable for someone to hold an agent morally blameworthy for some action without that person having any expectation that the agent not have performed the action as long as a reasonable expectation *could* be had. Consider the following situation. Suppose Deborah holds Raymond morally blameworthy for eating the last slice of chocolate cake even though she didn't expect him to refrain from eating the slice. On the view being considered, Deborah reasonably holds Raymond blameworthy as long as a morally reasonable expectation *could* be had that Raymond did otherwise. I find this implausible. Why should it matter whether she (or anyone else) could have reasonably expected Raymond to have done otherwise? Deborah did not have this expectation. If we assume that Deborah *could* have reasonably had the expectation that Raymond not eat the chocolate cake, what seems to follow is not that she *is* reasonable in holding Raymond blameworthy, but, rather, that she *could* have been reasonable in holding him blameworthy. Since she did not have this reasonable expectation, it does not seem reasonable for her to hold him blameworthy.

But Capes is not out of the woods. As he pointed out earlier, PAE\* (along with premise (W2)) does not entail PAP. However, PAE<sup>†</sup> (along with (W2')) does. And PAE<sup>†</sup> does not face the same difficulty that PAE\* does. PAE<sup>†</sup> concerns whether an agent *is* blameworthy. Clearly someone can be morally blameworthy for performing some action without anyone

actually having any expectation that she not have acted in that way. But it is far from clear that someone can be blameworthy without it being the case that someone who knew all the relevant facts *could* reasonably have expected her not to have acted in the way she did. To show this, Capes needs some other argument, one he does not provide. For this reason Capes has not shown that the key premise of Widerker's argument for PAP is false.

However, even if Capes has not demonstrated (or provided the materials to demonstrate) that  $PAE^\dagger$  is false, he raises a good question, one he puts forward in the passage quoted above. What reason is there to believe that  $PAE^\dagger$  is true? Without a good answer to this question, Widerker's argument for PAP is susceptible to the charge that it is undermotivated.

As I see the dialectic, we have two competing frameworks for moral responsibility. One is an alternative-possibilities framework; the other is an actual-sequence framework. The intuitions behind the W-defense and the L-reply that McKenna writes about are grounded in each of these frameworks, respectively. Frankfurt examples are intended to shift our thinking from an alternative-possibilities framework to an actual-sequence framework. They lead us to focus on what the agent did and why she did it, rather than on what else she could have done. The idea is that the actual-sequence framework is able to accommodate our intuitions regarding Frankfurt examples, while the alternative-possibilities framework is not. Thus, it will not do for Widerker to appeal to  $PAE^\dagger$  to support PAP. Both of these principles presuppose an alternative-possibilities framework. Someone who is unsure of which framework to adopt will not (and should not) be convinced by an appeal to  $PAE^\dagger$ . And those who have already adopted an actual-sequence framework should reject both  $PAE^\dagger$  and PAP. What Widerker needs to do is show us that there are important intuitions about

moral responsibility that an alternative-possibilities framework is able to capture that an actual-sequence framework is not able to.<sup>16</sup>

This is particularly important for Widerker to do because the Frankfurt defender and the PAP defender are not on equal footing here. Widerker has already granted (for the sake of discussion) that IRR-situations are possible. Thus, given the truth of IP\*, it seems possible that there are situations where the availability of alternative possibilities is irrelevant to whether an agent is morally responsible. That is, Frankfurt examples seem to provide us with good reason to favor the actual-sequence framework over the alternative-possibilities framework. This puts pressure on Widerker to provide compelling reasons for accepting PAE<sup>†</sup>. If he is unable to do so, then, contrary to what McKenna claims, we are not left with an “irresolvable tension” between intuitions. Rather, the actual-sequence framework comes out on top.

#### ***4.5. Justification for the Principle of Alternative Expectations***

Widerker (2005) offers three reasons for accepting PAE<sup>†</sup>.<sup>17</sup> First, it provides a natural explanation of why people are not blameworthy in situations where they were coerced or manipulated into performing the actions they did. Consider Bribe, for instance. Both Frankfurt defenders and PAP defenders agree that if Black activates his device, causing Jones to accept the bribe, then Jones is not blameworthy for the resulting action. But why not? Widerker thinks the explanation provided by PAE<sup>†</sup> is obvious. It would be unreasonable to expect Jones not to have performed the action he did.

---

<sup>16</sup> This point comes from comments Fischer made on an earlier draft of this chapter. He also makes the point in Fischer (2006a, pp. 209-210).

<sup>17</sup> The arguments Widerker offers are in support of PAE, but they apply equally to PAE<sup>†</sup>.

A second reason for accepting PAE<sup>†</sup> is that it provides a ready explanation for why an agent is sometimes morally blameworthy for failing to *try* to perform an action even if he is not blameworthy for failing to actually perform the act. For example, in Locke's locked room case, the man is not blameworthy for failing to leave the room (which he could not do since, unbeknownst to him, the door is locked) but he does seem to be blameworthy for not *trying* to leave the room. Why is he blameworthy for failing to *try* to leave the room but not blameworthy for failing to leave the room? The most natural answer, says Widerker, is that it is reasonable to expect the man to try to leave the room, but unreasonable to expect him to actually leave the room.

Widerker's third reason is that certain attitudes or sentiments that are associated with blame are mistaken if an agent can be morally blameworthy for performing an action he couldn't have avoided. Widerker offers remorse as an example. We believe that "[i]f S is aware that he acted culpably, it is appropriate for S to express remorse for his behavior" (Widerker, 2005, pp. 300, footnote removed). But this principle (Widerker calls it "RMS") is false if an agent can be blameworthy for something he couldn't have avoided. Widerker says that remorse involves a belief that one should not have acted as he did (which, in turn, involves an expectation that he not have done what he did). But if an agent became aware, after the fact, that he was in an IRR-situation in which he couldn't have done otherwise, it would be irrational for him to expect himself to have done otherwise, and therefore, inappropriate for him to feel remorse.

Palmer (2013a) provides an additional reason for accepting PAE<sup>†</sup>: it provides a natural explanation for a variety of situations where people are not blameworthy for their actions even though they could have done otherwise. He provides the following illustration. Suppose someone, Grace, flips a switch as she enters her hotel room, thinking it will turn on

the lights. What she doesn't know, however, is that the switch activates the fire alarm. Moreover, there is no reason for Grace to think that flipping the switch would trigger the fire alarm (there are no signs posted, the switch looks like a typical light switch, etc.). Intuitively, Grace is not blameworthy for activating the fire alarm. Palmer asks, *why* isn't Grace blameworthy? One natural explanation, he says, is that it would not be morally reasonable to expect Grace not to have flipped the switch. Why not? Because, given the circumstances, we would not think that she should have known this. Palmer points out that this explanation is an instance of PAE<sup>†</sup>: Grace is not morally blameworthy for flipping the switch because it would not be morally reasonable to have expected her not to have flipped the switch.

Palmer considers a possible objection to his argument; one that he thinks actually provides further support for PAE<sup>†</sup>. One might offer another way of explaining why Grace is not blameworthy, one that is consistent with the rejection of PAE<sup>†</sup>. For instance, the following epistemic condition on blameworthiness might be offered: An agent S is morally blameworthy for performing an action A only if (i) S believed that it would be wrong to A, and (ii) it is not the case that S should have believed it. Grace does not meet either of these conditions. Thus, one might argue that the most natural explanation of why Grace is not morally blameworthy is this epistemic principle. However, Palmer argues that even if this claim is granted, it *supports* PAE<sup>†</sup> rather than undermining it. How? Suppose the explanation for why Grace is not blameworthy is that she does not meet the epistemic conditions on blameworthiness. Palmer says that we should ask a broader question: Why should we accept these epistemic conditions on moral blameworthiness in the first place? He answers,

A natural answer to this question is to suppose that what explains why people are blameworthy for acting only if they believe that it would be morally wrong to act in that way (or, if they don't believe this, that they should do so) is that if it's the case that they *don't* believe their actions are wrong (and it's *not* true of them that they should believe this), then it's difficult to see how it could be morally reasonable, under the circumstances, to expect them not to have acted as they did. After all, if someone *doesn't* believe that acting in a particular way is wrong and it's not the case that he *should* think that this is wrong, then how can it be morally reasonable, given the circumstances, to expect him not to do the thing in question? (Palmer, 2013a, p. 562)

As Palmer sees it, a natural reason for accepting the epistemic conditions on moral blameworthiness is that it would be unreasonable to have expected someone who did not meet these conditions not to have done what she did.

What should we say about these four arguments for PAE<sup>†</sup>? I think the proponent of the actual-sequence approach to moral responsibility is able to account for our intuitions regarding blameworthiness in these cases in a way that is at least as plausible as PAE<sup>†</sup>. The actual-sequence approach forces us to look at the quality of will that is revealed by an agent's action, not what it is reasonable to have expected an agent to do. An agent's quality of will is revealed by what she does and why she does it. Thus, the Frankfurt defender might adopt what McKenna calls the "Quality of Will Thesis":

Being morally responsible and legitimately holding morally responsible are to be settled exclusively in terms of the moral quality of the will with which an agent acts. (McKenna, 2005, p. 172)

We can formulate a principle that rivals PAE<sup>†</sup> this way:

QWT: When an agent is non-derivatively morally blameworthy for an action, he is so in virtue of his having acted from a morally objectionable quality of will.<sup>18</sup>

McKenna's Quality of Will Thesis provides a natural explanation for the truth of IP\*. Why is it that facts that are irrelevant to a full explanation of what an agent did and why he did it have no bearing on an agent's moral responsibility? The answer provided by the Quality of Will Thesis is obvious. Such facts have no bearing on moral responsibility because they are irrelevant to assessing the quality of an agent's will. Frankfurt comes close to offering this reasoning himself. He says,

When a fact is in this way irrelevant to the problem of accounting for a person's action it seems quite gratuitous to assign it any weight in the assessment of his moral responsibility. Why should the fact be considered in reaching a moral judgment concerning the person when it does not help in any way to understand either what made him act as he did or what, in other circumstances, he might have done? (Frankfurt, 1969, p. 837)

The reason such facts play no part in assessing moral responsibility is that they do not reveal the quality of an agent's will. A person's quality of will is manifested "[w]hen a person acts for reasons of his own, and is guided entirely by his own beliefs and preferences" (Frankfurt, 2003, p. 340).

With QWT in hand, let us now consider each of the above cases. In the first case, QWT provides a natural explanation of why people are not blameworthy in situations where they were coerced or manipulated into performing the actions they did. Why isn't Jones morally

---

<sup>18</sup> Capes says that, "An agent has a morally objectionable quality of will just in case the beliefs, desires, values, intentions, plans, etc. in light of which the person acts exhibit a morally unjustifiable degree of ill will, indifference, or lack of due regard for others or for the moral considerations that bear upon her situation" (2012, p. 432).



blameworthy in a situation where Black activates his device? Because his accepting the bribe under those circumstances does not reveal the quality of his will (ill or otherwise). Now, consider the second case. QWT provides a plausible explanation of why an agent can be morally blameworthy for failing to *try* to perform an act even if he is not blameworthy for failing to perform the act. The man in the locked room is blameworthy for not trying to leave because in failing to *try* he reveals a morally ill will; and he is not blameworthy for failing to leave the locked room because this failure does not reveal a morally objectionable quality of will. In the fourth case, QWT explains why Grace is not blameworthy for flipping the switch that activated the fire alarm. She didn't know (nor should she have known) that the switch was connected to the fire alarm, and therefore did not reveal a morally ill will in flipping the switch.

Widerker's third argument is the best of the bunch. Let us grant that RMS is true. If remorse does in fact require an expectation that an agent not have done what he did, then Widerker has a compelling argument for PAE<sup>†</sup>. However, it is not clear why we should grant that remorse involves this expectation, and Widerker does not provide any argument for thinking that is does. In a footnote, he says that remorse differs from shame or regret. "Unlike in the case of remorse, one can feel shame or feel regret for what someone else did (say one's parent). Also, one can express both shame or regret for things one did not do voluntarily" (2005, pp. 300, footnote 331). However, it seems possible to capture this distinction between remorse and shame or regret without requiring that an agent has an expectation that he acted differently. Think about the case where Green does not go to work. His decision not to go to work is unaffected by his discovery that he is sick. Were he to feel remorse, it seems reasonable to think that his remorse would be grounded in the fact that his decision revealed an objectionable quality of will. He might wish that the *way* in

which he came to the decision not to go to work had been different even if he does not wish that the decision had been different. I see no reason why an agent in an IRR-situation, upon discovering that he was unable to do otherwise, couldn't feel remorse on the same grounds. He might wish that he had not performed the action on his own, but had been coerced, because of what it revealed about his will. This account of remorse that is grounded in QWT seems very plausible to me. Without some further explanation from Widerker, it is difficult to see why we should grant that remorse requires an expectation that the agent not have performed the action that he did.

It appears that QWT is able to provide an explanation of our intuitions in each of the above cases that is at least as natural as the explanation offered by  $PAE^\dagger$ . Thus, there seems to be no reason to prefer  $PAE^\dagger$  to QWT. Moreover, Widerker has not successfully demonstrated that there are important intuitions about moral responsibility that an alternative-possibilities framework is able to capture that an actual-sequence framework is unable to.

Given both  $IP^*$  and the possibility of IRR-situations, there seems to be good reason to reject the alternative-possibilities framework, along with  $PAE^\dagger$ , in favor of an actual-sequence framework. Widerker might resist this conclusion by denying the possibility of IRR-situations. However, were he to make this move, it would turn out that his argument does not circumvent the debate over the possibility of IRR-situations after all.  $PAE^\dagger$  itself (as well as the alternative-possibilities framework) would depend on IRR-situations being impossible. This is as it should be. As I see things, the debate over PAP really does hinge on whether IRR-situations are possible. And if my argument in Chapter 2 is successful, they are.

#### ***4.6. The Deontic Maxim Defense***

The second argument for PAP that I wish to consider involves Kant's intuitively plausible maxim that "ought" implies "can." Let us call this maxim the "Deontic Maxim." It can be formulated in this way:

DM: Necessarily, if an agent S ought to have done something other than A, then S can (is able) to do something other than A.

The argument was first introduced by Widerker (1991) and versions of the argument have been defended by David Copp (1997, 2003) and Daniel Speak (2005). Here is the argument (call it the "Deontic Maxim defense," or "DM-defense" for short):

- (D1) Necessarily, if an agent S is non-derivatively morally blameworthy for performing some action A, then it was objectively morally wrong for S to A.
- (D2) Necessarily, if it was objectively morally wrong for S to A, then S ought to have done something other than A.
- (D3) Necessarily, if S ought to have done something other than A, then S was able to do something other than A (DM).
- (D4) Therefore, necessarily, if S is non-derivatively morally blameworthy for A-ing, then S was able to do something other than A.

The conclusion is logically equivalent to PAP. If the DM-defense is successful, then we have a good reason for accepting PAP apart from PAP(CC). In the remainder of this chapter I shall consider several attempts to reject either premise (D1) or (D3) of the argument.<sup>19</sup>

---

<sup>19</sup> Gideon Yaffe (1999) has objected to premise (D2). An obligation not to perform some action is not equivalent to an obligation to perform some other action. Intuitively, one can meet an obligation not to perform some morally wrong act simply by refraining from performing the act. Therefore, (D2) seems to be false. Fischer, however, has argued that (D2) can be revised to avoid this worry, resulting in the following argument for PAP (2006c):

#### ***4.7. Premise (D1): The Objective View of Moral Blameworthiness***

Ishtiyaque Haji has labeled the first premise of the DM-defense the “Objective View” of moral blameworthiness (1998):

OV: Necessarily, if an agent S is non-derivatively morally blameworthy for performing some action A, then it was objectively morally wrong for S to A.<sup>20</sup>

Several cases have been offered to show that it is possible for an agent to be blameworthy for a certain action even though the action was not objectively morally wrong. Here is one from Haji. He provides a situation in which an agent performs an action intending harm, but the action ends up being beneficial (he calls it “Deadly’s Defeat”):

[D]oing the best she can to murder her patient, Dr. Deadly does what credible evidence to which she has access indicates will kill the patient—she injects the patient with medicine C. But...there has been an innocent error in diagnosis. Contrary to what Deadly believes, her patient is suffering from a disease that can be cured only by taking C. Giving C results in the lucky patient’s full recovery. Although it is arguably

- 
- (D1) Necessarily, if an agent S is non-derivatively morally blameworthy for performing some action A, then it was objectively morally wrong for S to A.  
(D2') Necessarily, if it was objectively morally wrong for S to A, then S ought to have not-Aed.  
(D3') Necessarily, if S ought to have not-Aed, then S was able to have not-Aed (from DM).  
(D4') Therefore, necessarily, if S is non-derivatively morally blameworthy for A-ing, then S was able to have not-Aed.

What I have to say regarding the DM-defense does not hinge on this distinction. The real action, I believe, involves premises (1) and (3). So I shall set this worry aside. The reader is welcome to substitute this revised version of the argument for the one I use throughout the remainder of this chapter.

<sup>20</sup> It is objectively morally wrong for an agent S to perform some action A iff, all things considered, S was morally required not to A.

obligatory for Deadly to inject her patient with C, it appears that Deadly is to blame for the injection of this drug.<sup>21</sup> (Haji, 1998, p. 146)

Haji thinks Deadly is morally blameworthy for administering C, but contrary to OV, it was not objectively wrong for her to do so since administering C was obligatory.

Haji considers and responds to one way of objecting to Deadly's Defeat as a counterexample to OV. He says that the proponent of OV might grant that it is not objectively wrong for Deadly to administer C, but argue that Deadly is not blameworthy for performing this action since administering C is obligatory. The OV theorist might go on to explain that this does not mean that Deadly is not blameworthy for *something*. He might say that what Deadly *is* blameworthy for is for *trying (or attempting)* to kill the patient, and this is something that it is objectively morally wrong for her to do.

Haji finds this way of responding unpersuasive. He says that it "appears to rest on the dubious principle that it is always the case that if one attempts to do something, and that thing is wrong, then one's attempt to do that thing is also wrong" (Haji, 1998, pp. 146, footnote omitted).<sup>22</sup> Why is this principle dubious? Because it would be wrong for Deadly to kill the patient, yet "Deadly's attempt to kill the patient (where the attempt consists, at least partly, in Deadly's injecting the patient with medicine C) does not seem to be wrong"

---

<sup>21</sup> It would be misleading to say that Deadly *cured* the patient. The way that Haji phrases this is just right: Deadly did something (giving the patient C) that resulted in the patient's full recovery.

<sup>22</sup> Dana Nelkin offers a few alternative principles: (i) if one attempts to do something that is aimed at producing harm that serves no good purpose, then one's attempt to do that thing is wrong; (ii) if one attempts to do something that he should have known was likely to produce harm for no good purpose, then one's attempt to do that thing is wrong (Nelkin, 2011, p. 107).

(Haji, 1998, p. 146). But why should the proponent of OV accept Haji's claim that Deadly's *attempt* to kill the patient is not objectively wrong?<sup>23</sup>

Haji defends this claim by arguing that Deadly's attempt to kill the patient is obligatory, and therefore, not objectively morally wrong. He says,

[I]f we grant that injecting the drug is obligatory for Deadly, then, if Deadly's attempt to kill the patient consists (at least partly) in Deadly's injecting the patient with C, the attempt itself (given the reasonable assumption that Deadly was able to refrain from attempting to inject the patient), must itself be obligatory, on the assumption that this "prerequisites" deontic principle is correct: if S cannot do A without doing B (out of, for example, logical or physical necessity), and S can refrain from doing B, then, if S ought to do A, then S ought to do B. As in Deadly's case, it seems that Deadly cannot inject

---

<sup>23</sup> What Haji might be thinking is that there is only one action that Deadly performs, and therefore, this one action is not objectively wrong since performing the action is obligatory. That is, Haji might have a coarse-grained theory of action individuation in mind. However, Capes (2012, pp. 422-423) has recently explained that the proponent of OV could reasonably argue that Deadly's attempt to kill the patient is morally wrong on both a fine-grained theory of action individuation and a coarse-grained theory.

On the fine-grained approach, Deadly has performed several actions: injecting the patient with medicine C, curing the patient, and attempting to kill the patient. However, the OV theorist can claim that the only action Deadly is morally blameworthy for is for attempting to kill the patient, but this is in accord with OV since it was objectively morally wrong for her to do so. Concerning the other actions for which it is not objectively morally wrong for Deadly to perform (in particular, injecting the patient with medicine C), she is not blameworthy for performing them.

On the coarse-grained approach, there is just one action that Deadly performed. However, whether she is blameworthy for that action will depend, in part, on how her action is described. Capes notes that "when actions are individuated coarsely, it is actions under descriptions that are the targets of various act-evaluations" (Capes, 2012, p. 422). Thus, under some descriptions Deadly's action will not be objectively morally wrong. For example, under the description *injecting the patient with medicine C*, her action is not objectively wrong and it is also not something for which she is morally blameworthy. Under the description *attempting to kill the patient*, the OV theorist can claim that although Deadly's action was morally blameworthy, contrary to Haji's claim, it was also objectively wrong.

the patient without attempting to inject the patient, the attempt must be obligatory if the injection itself is. (Haji, 1998, pp. 146, footnote omitted)

Capes objects. He says that the OV theorist could grant both that injecting C is obligatory (and therefore, not objectively wrong) and that Haji's deontic principle is correct, yet still resist Haji's conclusion. Consider the result of applying the deontic principle to Haji's case:

If Deadly cannot *inject the patient with C* without *attempting to kill the patient*, and Deadly can refrain from *attempting to kill the patient*, then if Deadly ought to *inject the patient with C*, then Deadly ought to *attempt to kill the patient*.

To get the conclusion that Deadly ought to attempt to kill the patient, the antecedent must be satisfied. However, Capes asks why we should think that it is. Why couldn't Deadly inject the patient with C without attempting to kill the patient? Certainly it is not logically or physically impossible. Capes continues,

To reach that conclusion Haji would need the following claim to be true: Deadly cannot (intentionally) inject the patient with C without trying to kill him. However, I see no reason whatsoever to accept that claim and, moreover, it seems straightforwardly false. (Capes, 2012, p. 424)

Capes is surely correct to say that Deadly *could* inject the patient with C without trying to kill him. However, much like Needed Medication (from Chapter 3) in which it is within Daphne's voluntary control to give Tal whatever is in the jar marked 'aspirin,' such an option is not deliberatively significant. Given that Deadly has no idea that injecting medicine C would bring about the patient's recovery, it would be unreasonable for her to do so, even if she wanted to cure the patient instead of killing him. Thus, it seems that Capes' response could be avoided by revising Haji's deontic principle:

If S's doing A without doing B is not *deliberatively significant* for S given her circumstances, and S can refrain from doing B, then if S ought to do A, then S ought to do B.

This, I think, captures the essence of Haji's reasoning. Clearly Deadly has no reason to inject the patient with C without trying to kill him. If the revised deontic principle is correct, and injecting C is obligatory, then it follows that Deadly's attempt to kill the patient is obligatory.

The trouble, as I see it, with Haji's argument is that it is unclear why we should grant this revised deontic principle. It's not difficult to imagine a counterexample. Consider the following. Doctor Grim is testing how effective a new drug D is at killing people. During her hospital rounds she injects all of her patients with D. However, contrary to what Grim (non-culpably) believes, one patient is suffering from a disease that can be cured only by receiving D. Grim's administering D results in the deaths of all of her patients except the lucky one who is cured by receiving D. Given her circumstances, Grim's administering D to just the lucky patient without administering it to all of her patients is not *deliberatively significant* for her. And Grim can refrain from administering D to all of her patients. However, although it is arguably objectively morally obligatory for Grim to administer D to the one patient it would cure, it does not follow that administering D to all of the patients is objectively morally obligatory. Indeed, doing so is objectively morally wrong. Therefore, the revised deontic principle is false.

Returning to Deadly's Defeat, the OV theorist need not accept Haji's claim that Deadly is objectively morally obligated to try (or attempt) to kill the patient. Thus, it is still open for the OV theorist to claim that despite the fact that it is not objectively morally wrong for



Deadly to administer C, she is not morally blameworthy for administering C, but rather, for attempting to kill the patient.

Another way of objecting to this response from the OV theorist comes from Michael Zimmerman. He provides a reason for thinking that Haji is correct in claiming that Deadly *is* morally blameworthy for injecting the patient with C. He says that “acting freely in the belief that one is doing objective moral wrong is sufficient for being morally blameworthy for one’s behavior (but not sufficient for actually doing objective moral wrong)” (M. J. Zimmerman, 1997, p. 234). Call this principle “Belief Entails Blameworthiness” (or “BEB” for short). If Zimmerman is correct, and if we assume that Deadly believes it is objectively morally wrong to administer C to the patient, then it follows that she *is* blameworthy for administering C to the patient.

However, there appear to be counterexamples to BEB. Consider the case of Huckleberry Finn who believed that rescuing Jim, a slave, from the Phelps’ farm was objectively morally wrong but did it anyway. If BEB is true then counterintuitively Huckleberry Finn is morally blameworthy for rescuing Jim.<sup>24</sup> Capes has recently offered another counterexample that I find persuasive:

It’s 1943, and Corrie is hiding Jewish refugees in her home in Haarlem, Holland. One day a group of German soldiers ask her if she is harboring any Jews. Now Corrie takes

---

<sup>24</sup> Zimmerman uses the Huckleberry Finn case in the same article to argue that a related principle is false (call it “Belief Entails Wrongdoing”):

BEW: Necessarily, if an agent S believes that performing an action A is objectively morally wrong, then it is objectively morally wrong for S to perform A.

Zimmerman says that if we accept BEW, we arrive at the counterintuitive conclusion that it is objectively morally wrong for Huckleberry Finn to rescue Jim. Zimmerman asks, “If one can be mistaken in *failing* to believe that one is doing wrong, why can one not be mistaken (by virtue of an overly sensitive conscience, perhaps, or simply by virtue of a misreading of one’s situation) in *believing* that one is doing wrong?” (M. J. Zimmerman, 1997, p. 233). Intuitively, Huckleberry Finn is mistaken in believing that it is wrong for him to rescue Jim. But if the above principle is true, he cannot be mistaken about this belief.

the Ten Commandments very seriously and, like Kant, believes that lying is always morally impermissible, no matter what. However, she also believes that it would be objectively wrong to tell the soldiers the truth, for by doing so she would effectively be handing over the refugees to be tortured and killed, which, of course, would be wrong. She doesn't regard remaining silent as morally permissible either, for if she keeps silent, she believes that the officers will get suspicious, raid the house, find the refugees and whisk them all, herself and her family included, off to concentration camps where they will almost certainly perish. In short, Corrie believes that every course of action available to her is objectively morally wrong. In the end, Corrie freely lies to the soldiers, despite believing that she is morally obliged not to lie to them. The reason she does this, though, is to prevent herself, her family and those she is hiding from being unjustly imprisoned and killed. If there had been a course of action available to her that she regarded as morally permissible, she most certainly would have opted for it. (Capes, 2012, p. 425)

Intuitively, Corrie is not blameworthy for lying to the soldiers even though she freely lied to them in the belief that in doing so she was doing something objectively morally wrong. If this is correct, then BEB is false. And if BEB is false, then it is not clear why the OV theorist should grant that Deadly is morally blameworthy for administering C to the patient. Therefore, the OV theorist's reply to Haji's Deadly's Defeat case stands.

However, there is another putative counterexample to OV that I think is successful. It is offered by Capes, and it does not depend on the truth of BEB. He calls it "Kill Bill":

Beatrix freely shot and killed Bill. She did this despite believing that it was objectively wrong to kill Bill, that it was within her power to avoid killing him and, indeed, that it was within her power to avoid wrongdoing altogether. Why, then, did Beatrix kill Bill?

What motivated her to perform this action? A significant role was played by her hatred of Bill and her (no doubt morally unjustified) desire to rid the world of him.

Unbeknownst to Beatrix, however, Bill was just about to torture and kill her daughter, and the only way she could have prevented him from doing so was to shoot and kill him.

(Capes, 2012, pp. 428-429, footnote omitted)

Why think Beatrix is blameworthy for killing Bill? Instead of appealing to BEB, Capes says,

[A] person is blameworthy for her conduct just in case the behavior in question serves as a basis for a certain kind of negative moral judgment or appraisal of the person, at least with respect to that particular episode in her life...Someone who freely *A*-ed as a result of her ill will and lack of due regard for the wellbeing of another, despite believing that it was objectively wrong to *A* and that it was within her power to avoid wrongdoing entirely, arguably merits moral criticism for *A*-ing. (Capes, 2012, p. 429)

Beatrix certainly seems to merit moral criticism for killing Bill. She performed the action freely as a result of her ill will and lack of due regard for the wellbeing of Bill, and she did so despite believing it was objectively wrong and that it was within her power to refrain from the act.

Let us grant that Beatrix is morally blameworthy for killing Bill. Was it objectively morally wrong for Beatrix to kill Bill? It seems not. Why not? Capes says it follows from a plausible principle that governs killing in self-defense and in defense of the innocent, the Principle of Self-Defense (PSD):

[I]t is morally permissible to kill an unjust aggressor, if doing so is the only way to save your own life or the lives of those you have a duty to protect, provided that you can do

so without doing something else that you are morally required not to do. (Capes, 2012, p. 430)

Capes defines an *unjust aggressor* as someone who, without good reason, is trying to kill you or those you have a duty to protect and is blameworthy for doing so. If we assume that Bill was an unjust aggressor, and we assume that killing Bill did not require Beatrix to do anything that she was morally required not to do, then Beatrix's killing Bill is morally permissible, and therefore, not objectively morally wrong. Capes concludes that Kill Bill is a counterexample to OV: Beatrix is non-derivatively morally blameworthy for killing Bill even though doing so was not objectively morally wrong.

This is a good case, and I think Capes is correct to reject OV. However, even if Capes is mistaken and OV can be successfully defended against putative counterexamples, the DM-defense fails. This is because if one accepts OV, one should reject DM.

#### ***4.8. Premise (D3): The Deontic Maxim***

Before considering objections to DM, I wish to look at two ways of motivating the maxim offered by David Copp (1997, 2003, 2008). I shall argue that both of these attempts fail, which will put us in a position to better appreciate the objections to DM.

Copp's first argument is an argument from fairness. He begins,

The most basic motivation for [DM], it seems to me, begins with the thought that it would be unfair to expect a person to do something, or to demand or require that she do it, if she lacked the ability to do it. This thought is about what we might call 'agent-requirements,' which arise in cases in which an authoritative agent requires someone under her authority or jurisdiction to do something. An example might be a situation in which a boss requires an employee to do something that the employee lacks the ability to do. A supervisor at the post office might demand that a mail carrier cook a soufflé for

everyone in the post office in the next five minutes when the mail carrier does not even know what a soufflé is...The intuition is that agent requirements of this kind are morally unfair when the person of whom the demand is made lacks the ability to comply. (Copp, 2003, p. 271).

From this intuition about *agent* requirements, Copp thinks we can draw a similar intuition about *moral* requirements. Morality would be unfair if it demanded of a person that she do something she is unable to do. Why think this? Copp says that the intuition about agent requirements is in tension with the idea that someone could be morally required to do something but unable to do it: “[I]f there would be no unfairness in the fact that a person is morally required to do something she is unable to do, then it is unclear why it would be unfair, or where the unfairness would be, in an agent’s requiring a person to do something who is unable to do it” (Copp, 2003, p. 272). But, of course, “it is not intelligible to suppose that morality itself might be morally unfair” (Copp, 2003, p. 274). Therefore, it must not be the case that someone could be morally required to do something she is unable to do. That is, DM is true.

Copp’s second argument is an argument from the relevance of moral requirements to action guidance. He says,

Moral requirements are action-guiding. That is, their point is to guide agents’ decisions among their alternatives. All-in requirements partition an agent’s alternatives at the point of action into those that are permissible and those that are not. Hence, knowledge of what one is all-in morally required to do enables one to select a permissible action from among one’s alternatives. Given this, an adequate theory would imply that an agent is all-in morally required to do A only if she can do A. Otherwise, a permissible action would not be among her alternatives, since a person’s alternatives at a time are the

things she can do at that time. This reasoning applies as well to requirements not to do things. For if a person might be all-in required not to do A even if she could not avoid doing A, a permissible action might not be among her alternatives. But then an adequate theory would imply that if an agent is all-in morally required not to do A, she can do something else. (Copp, 2008, p. 71)

If an agent could be morally required to perform an action she is unable to perform, then moral requirements are not relevant (in a characteristic way) to the decisions agents make. But moral requirements *are* relevant, says Copp. Therefore, it must *not* be the case that an agent could be morally required to perform an action she is unable to perform. That is, DM is true.

Let us begin with Copp's argument from fairness. Rob van Someren Greve (2014) has recently raised two objections to the argument, both of which are persuasive. First, he says that claiming that morality can be fair or unfair involves a category mistake:

[W]hat can be fair or unfair is what we expect of others, how we treat them, how we distribute goods among them, and so on, but ascribing fairness or unfairness to *deontic facts* (i.e., an action's being obligatory or wrong) looks simply incoherent—deontic facts or properties are just not the sort of thing that can be fair or unfair. (van Someren Greve, 2014, p. 918)

Thus, the reason it would be unintelligible to say that moral requirements could be morally unfair is not because moral requirements must be fair, but because moral requirements are not the sorts of things to which moral fairness or unfairness can be properly applied. It would be just as unintelligible to say that moral requirements could be fair.<sup>25</sup>

---

<sup>25</sup> Nomy Arpaly puts it this way:

Fairness is part of morality, and so to talk of morality being fair is like talking about morality being moral (or aesthetics being beautiful, say). If morality is fair, does it mean

This seems correct. Copp's application of our intuitions about the fairness of agent requirements to moral requirements is inappropriate. Part of the way he motivates this move is by talking about morality *making demands* of agents. But the intuitive pull of this sort of talk seems to come through a personification of morality, which is misleading. For this reason, it is better to use agent-neutral language. When we talk about whether moral requirements can be fair or unfair, instead of whether the demands of morality can be fair or unfair, this intuition is not nearly so strong. The answer, then, to Copp's challenge—if it is not unfair for a person to be morally required to do something she is unable to do, why would it be unfair for an agent to require a person to do something she is unable to do?—is that, unlike agent requirements, moral requirements are not the sorts of things that can be fair or unfair.

However, even if sense can be made of saying that moral requirements must be morally fair, Copp's argument still fails. The argument makes the false assumption that moral requirements are never unfair. But moral requirements are sometimes unfair. Consider the following case offered by van Someren Greve. Imagine Sally, Jimmy, and Billy are all lifeguards at the community pool for the summer. Throughout the summer Jimmy and Billy just sat around while Sally did all the work of saving neighborhood children from drowning. Today is no different. A child is about to drown and Sally knows that Jimmy and Billy won't do anything to help. Clearly Sally is morally required to rescue the child. But it also seems to be unfair that she is, since she is the only lifeguard of the three who has been doing her job. As van Someren Greve points out,

---

that I can say, "Look, morality, I have done three superarrogatory acts last week. Surely I can get away with lying to a student today." This is where Anscombe would probably suspect us of being under the spell of regarding morality as a set of commands from a celestial boss, who can be fair or unfair to us. (Arpaly, 2006, p. 107)

[I]t is often unfair that the burden of doing something falls on a specific person's shoulders, solely because others who could have done their part did not in fact do so. Unfair as it may be, it seems unwarranted to conclude that therefore these people are under no obligation to act, provided there is enough at stake. (van Someren Greve, 2014, p. 919)

Let's turn to Copp's second argument for DM. It centers around the claim that moral requirements would be "pointless" if they could not guide an agent's deliberation. Consider a case offered by Copp. Imagine that Madeleine is considering whether or not to lie. However, unbeknownst to Madeleine, Count Vener is able to see into Madeleine's mind and control it if he chooses. He wants Madeleine to lie. If he sees that she is forming the desire not to lie, he will immediately intervene, making it the case that she desires to lie. As things happen, Madeleine does lie, without any intervention from the count. Was Madeleine morally required not to lie? Copp says that she was not. As he sees it, if Madeleine was morally required not to lie, then this moral requirement would be irrelevant to her deliberative process since not lying is not an option for her.

I agree with Copp that moral requirements play an important role in guiding our practical deliberation. The problem with his argument, as both Arpaly (2006, p. 108) and Fischer (2006c, pp. 206-207) have pointed out, is that the set of alternatives one *believes* that one has can be larger than the set of alternatives that are *actually* available. But it is the alternatives that one believes one has that are relevant to one's practical deliberation. I can deliberate over whether to eat a slice of pizza from the fridge for lunch even if, unbeknownst to me, my wife has already eaten it. A crucial feature of Copp's case is that Madeleine does not know that the presence of Count Vener makes it impossible for her to tell the truth. Thus, the moral requirement *do not lie* can still guide her practical deliberation since she is



unaware of the circumstances that prevent her from telling the truth. As Fischer notes, “Moral requirements insert themselves into the space of epistemic possibilities, not directly into the space of metaphysical possibilities” (Fischer, 2006c, p. 207).

So much for Copp’s attempts at motivating DM. I shall conclude this section with two arguments against DM. The first rests on Frankfurt examples involving omissions.<sup>26</sup> These are examples in which an agent is morally responsible for failing to perform some action even though he could not have performed the action. These examples can be used as direct counterexamples to DM. Recall the case of Sally the lifeguard. Suppose Sally is standing at the edge of the pool and she sees a small child drowning. All she has to do is reach her hand out for the child to grab, but she does not do so. Unknown to Sally, Black has implanted a neural device that happens to block every neural pathway except the one Sally is actually using. But the device plays no role in Sally’s failure to reach her hand out.<sup>27</sup> Intuitively, she didn’t do what she ought to have done. Sally had a moral obligation to reach her hand out to rescue the child even though she was unable to do so.<sup>28</sup>

Of course, the proponent of DM may not share my intuitions regarding this case, so I offer a second argument. Recall that PAP(CC) grounds the *prima facie* plausibility of PAP. If one rejects PAP(CC), absent other arguments in support of PAP, one has little reason to

---

<sup>26</sup> Sinnott-Armstrong (1988) rejects the maxim “ought implies can” for reasons independent of issues raised in this dissertation. He argues that we should give up the maxim in light of the possibility of moral dilemmas.

<sup>27</sup> Fischer (2006b, pp. 24-25) offers a similar case. However, he argues that Sally is morally responsible.

<sup>28</sup> Importantly, this argument does not beg the question against incompatibilists or PAP defenders. Given that agents may have *excuses* for wrongdoing—that is, an agent may perform an action that is objectively morally wrong without being blameworthy for performing that action (see J. L. Austin, 1956)—the rejection of DM is compatible with the position that causal determinism rules out moral blameworthiness. Note that Widerker, an incompatibilist and proponent of the W-defense for PAP, rejects DM (2005, pp. 303-304).

accept it. A similar move can be made regarding DM. As stated, DM does not require a deep conceptual connection between moral obligation and alternative possibilities, but only that in every possible world in which one is morally obligated one also happens to have alternative possibilities. The intuitive principle that grounds an alternative possibilities requirement on moral obligation is the following:

DM(CC): When an agent was morally obligated to have done something other than A, he was so partly in virtue of his then having been able to have done otherwise.

If one rejects DM(CC), then one has little reason to accept DM. An argument for PAP(CC) using DM(CC) goes as follows:

(D1") When an agent is non-derivatively morally blameworthy for performing some action, he is so partly in virtue of the fact that it was objectively morally wrong for him to perform that action (call this "OV(CC)").

(D2) Necessarily, if it is objectively morally wrong for an agent to perform an action, then the agent is morally obligated to have done otherwise.

(D3") When an agent is morally obligated to have done otherwise, he is so partly in virtue of having been able to have done otherwise (DM(CC)).

(D4") Therefore, when an agent is non-derivatively morally blameworthy for performing some action, he is so partly in virtue of having been able to do otherwise (PAP(CC)).

PAP(CC) is false. So, given the truth of both (D1") and (D2), it follows that DM(CC) is false. The proponent of DM might object to (D1"). However, (D1") grounds the intuitive plausibility of OV. Without (D1"), it is difficult to see why one should accept OV. Of course, one could reject OV, but then the DM-defense for PAP would fail. What about (D2)? (D2) is a premise that also shows up in the DM-defense. So, if the proponent of DM

rejects (D2), then these DM-defense also fails. Thus, if one accepts both (D1") and (D2), then there is little reason to accept DM.

#### ***4.9. Concluding Remarks***

In this chapter I considered two arguments in favor of PAP. One turns on the Principle of Alternative Expectations and the other requires both the “objective view” of moral blameworthiness and the maxim “ought” implies “can.” Some have thought that rejecting these principles comes at too steep a price. But I have argued that this is not the case. Our intuitions about moral blameworthiness can be preserved even if these principles are false. Moreover, there are good reasons to reject them.

If my argument from Chapter 2 is correct, it is possible that there are situations where the availability of alternative possibilities is explanatorily irrelevant to an agent’s moral responsibility. Thus, Frankfurt examples give us good reason to shift our thinking from an alternative-possibilities framework for moral responsibility to an actual-sequence framework. Moreover, as we saw in this chapter, making this shift does not appear to require us to give up any important intuitions about moral responsibility that can be accounted for by an actual-sequence framework. If this is correct, then we should be skeptical of the truth of PAP. For all we know, it *could* be true, but we have little reason to believe that it is.

## References

- Arpaly, Nomy. (2006). *Merit, Meaning, and Human Bondage: An Essay on Free Will*. Princeton: Princeton University Press.
- Austin, J. L. (1956). A Plea for Excuses. *Proceedings of the Aristotelian Society*, 57, 1-30.
- Berofsky, Bernard. (2002). Ifs, Cans, and Free Will: The Issues. In R. H. Kane (Ed.), *The Oxford Handbook of Free Will* (pp. 181-201). Oxford: Oxford University Press.
- Campbell, Joseph Keim. (2008). New Essays on the Metaphysics of Moral Responsibility. *Journal of Ethics*, 12(3/4), 193-201.
- Campbell, Joseph Keim. (2011). *Free Will*. Cambridge: Polity Press.
- Campbell, Joseph Keim, O'Rourke, Michael, & Shier, David. (2004). *Freedom and Determinism*. Cambridge: MIT Press.
- Capes, Justin A. (2010). The W-defense. *Philosophical Studies*, 150, 61-77.
- Capes, Justin A. (2012). Blameworthiness Without Wrongdoing. *Pacific Philosophical Quarterly*, 93(3), 417-437.
- Chisholm, Roderick. (1964). Human Freedom and the Self. In R. Kane (Ed.), *Free Will* (pp. 47-58). Cambridge: Blackwell.
- Clarke, Randolph. (2003). *Libertarian Accounts of Free Will*. Oxford: Oxford University Press.
- Clarke, Randolph. (2010). Intentional Omissions. *Noûs*, 44(1), 158-177.
- Clarke, Randolph. (2014). *Omissions: Agency, Metaphysics, and Responsibility*. Oxford University Press.
- Coates, D. Justin, & Tognazzini, Neal A. (Eds.). (2013). *Blame: Its Nature and Norms*. New York: Oxford University Press.
- Cohen, Daniel, & Handfield, Toby. (2007). Finking Frankfurt. *Philosophical Studies*, 135(3), 363-374.
- Copp, David. (1997). Defending the Principle of Alternate Possibilities: Blameworthiness and Moral Responsibility. *Noûs*, 31(4), 441-456.
- Copp, David. (2003). 'Ought' Implies 'Can,' Blameworthiness, and the Principle of Alternative Possibilities. In D. Widerker & M. McKenna (Eds.), *Moral*

*Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities* (pp. 265-300). Aldershot: Ashgate.

- Copp, David (Ed.). (2006). *The Oxford Handbook of Ethical Theory*. Oxford: Oxford University Press.
- Copp, David. (2008). 'Ought' Implies 'Can' and the Derivation of the Principle of Alternate Possibilities. *Analysis*, 68(297), 67–75.
- Della Rocca, Michael. (1998). Frankfurt, Fischer and Flickers. *Noûs*, 32(1), 99-105.
- Dennett, Daniel C. (1984). *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge: MIT Press.
- Ehring, Douglas. (1997). *Causation and Persistence: A Theory of Causation*. Oxford: Oxford University Press.
- Ekstrom, Laura W. (1998). Protecting Incompatibilist Freedom. *American Philosophical Quarterly*, 35(3), 281-291.
- Ekstrom, Laura W. (2002). Libertarianism and Frankfurt-Style Cases. In R. Kane (Ed.), *The Oxford Handbook of Free Will* (pp. 309-322). Oxford: Oxford University Press.
- Elzein, Nadine. (2013). Pereboom's Frankfurt Case and Derivative Culpability. *Philosophical Studies*, 166(3), 553-573.
- Fara, Michael. (2008). Masked Abilities and Compatibilism. *Mind*, 117(468), 843-865.
- Fischer, John, Kane, Robert, Pereboom, Derk, & Vargas, Manuel. (2007). *Four Views on Free Will*. Cambridge: Blackwell.
- Fischer, John Martin. (1982). Responsibility and Control. *The Journal of Philosophy*, 79(1), 24-40.
- Fischer, John Martin. (1994). *The Metaphysics of Free Will: An Essay on Control*. Cambridge: Blackwell.
- Fischer, John Martin. (1999). Recent Work on Moral Responsibility. *Ethics*, 110(1), 93–139.
- Fischer, John Martin. (2002). Frankfurt-Type Examples and Semi-Compatibilism *The Oxford Handbook of Free Will*: Oxford University Press.
- Fischer, John Martin. (2003). Responsibility and Alternative Possibilities. In D. Widerker & M. McKenna (Eds.), *Freedom, Responsibility, and Agency: Essays on the Importance of Alternative Possibilities* (pp. 27-52). Aldershot: Ashgate.
- Fischer, John Martin. (2006a). Free Will and Moral Responsibility *My Way: Essays on Moral Responsibility* (pp. 182-216). Oxford: Oxford University Press.

- Fischer, John Martin. (2006b). *My way: Essays on Moral Responsibility*. New York: Oxford University Press.
- Fischer, John Martin. (2006c). "Ought-Implies-Can," Causal Determinism, and Moral Responsibility *My Way: Essays on Moral Responsibility* (pp. 217-222). Oxford: Oxford University Press.
- Fischer, John Martin. (2010). The Frankfurt Cases: The Moral of the Stories. *Philosophical Review*, 119(3), 315-336.
- Fischer, John Martin. (2013). The Deterministic Horn of the Dilemma Defence: A Reply to Widerker and Goetz. *Analysis*, 73(3), 489-496.
- Fischer, John Martin, & Ravizza, Mark. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press.
- Frankfurt, Harry G. (1969). Alternate Possibilities and Moral Responsibility. *Journal of Philosophy*, 66(3), 829-839.
- Frankfurt, Harry G. (1971). Freedom of the Will and the Concept of a Person. *Journal of Philosophy*, 68(1), 5-20.
- Frankfurt, Harry G. (2003). Some Thoughts Concerning PAP. In D. Widerker & M. McKenna (Eds.), *Freedom, Responsibility, and Agency: Essays on the Importance of Alternative Possibilities* (pp. 339-348). Aldershot: Ashgate.
- Franklin, Christopher Evan. (2011a). Farewell to the Luck (and Mind) Argument. *Philosophical Studies*, 156(2), 199-230.
- Franklin, Christopher Evan. (2011b). Neo-Frankfurtians and Buffer Cases: The New Challenge to the Principle of Alternative Possibilities. *Philosophical Studies*, 152, 189-207.
- Ginet, Carl. (1966). Might We Have No Choice? In K. Lehrer (Ed.), *Freedom and Determinism* (Vol. 7, pp. 87-104). New York: Random House.
- Ginet, Carl. (1996). In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing. *Philosophical Perspectives*, 10, 403-417.
- Ginet, Carl. (2002). Book Review. Living Without Free Will. Derk Pereboom. *Journal of Ethics*, 6(3), 305-309.
- Ginet, Carl. (2003). In Defense of the Principle of Alternative Possibilities: Why I don't Find Frankfurt's Argument Convincing. In D. Widerker & M. McKenna (Eds.), *Freedom, Responsibility, and Agency: Essays on the Importance of Alternative Possibilities* (pp. 75-90). Aldershot: Ashgate.
- Ginet, Carl, & Palmer, David. (2010). On Mele and Robb's Indeterministic Frankfurt-Style Case. *Philosophy and Phenomenological Research*, 80(2), 440-446.

- Goetz, Stewart. (2005). Frankfurt-Style Counterexamples and Begging the Question. *Midwest Studies in Philosophy*, 29(1), 83-105.
- Haji, Ishtiyaque. (1998). *Moral Appraisability: Puzzles, Proposals, and Perplexities*: Oxford University Press.
- Haji, Ishtiyaque, & McKenna, Michael. (2004). Dialectical Delicacies in the Debate about Freedom and Alternative Possibilities. *The Journal of Philosophy*, 101(6), 299-314.
- Haji, Ishtiyaque, & McKenna, Michael. (2006). Defending Frankfurt's Argument in Deterministic Contexts: A Reply to Palmer. *The Journal of Philosophy*, 103(7), 363-372.
- Hunt, David. (2000). Moral Responsibility and Unavoidable Action. *Philosophical Studies*, 97(2), 195-227.
- Hunt, David. (2003). Freedom, Foreknowledge, and Frankfurt. In D. Widerker & M. McKenna (Eds.), *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities* (pp. 159-183). Aldershot: Ashgate.
- Hunt, David. (2005). Moral Responsibility and Buffered Alternatives. *Midwest Studies in Philosophy*, 29(1), 126-145.
- Hunt, David, & Shabo, Seth. (2013). Frankfurt Cases and the (In)significance of Timing: A Defense of the Buffering Strategy. *Philosophical Studies*, 164(3), 599-622.
- Kane, Robert. (1996). *The Significance of Free Will*. Oxford: Oxford University Press.
- Kane, Robert. (2000). Responses to Bernard Berofsky, John Martin Fischer and Galen Strawson. *Philosophy and Phenomenological Research*, 60(1), 157-167.
- Kane, Robert (Ed.). (2002). *The Oxford Handbook of Free Will*. Oxford: Oxford University Press.
- Kane, Robert. (2003). Responsibility, Indeterminism and Frankfurt-style Cases: A Reply to Mele and Robb. In D. Widerker & M. McKenna (Eds.), *Freedom, Responsibility, and Agency: Essays on the Importance of Alternative Possibilities* (pp. 91-106). Aldershot: Ashgate.
- Kane, Robert. (2004). Agency, Responsibility, and Indeterminism: Reflections on Libertarian Theories of Free Will. In J. K. Campbell, M. O'Rourke & D. Shier (Eds.), *Freedom and Determinism* (pp. 70-88). Cambridge: MIT Press.
- Kane, Robert. (2005). *A Contemporary Introduction to Free Will*. New York: Oxford University Press.
- Lehrer, Keith. (1968). Cans Without Ifs. *Analysis*, 29(1), 29 - 32.

- Leon, Felipe, & Tognazzini, Neal. (2010). Why Frankfurt-Examples Don't Need to Succeed to Succeed. *Philosophy and Phenomenological Research*, 80(3), 551-565.
- Levy, Neil, & McKenna, Michael. (2009). Recent Work on Free Will and Moral Responsibility. *Philosophy Compass*, 4(1), 96-133.
- Locke, John. (1979). *An Essay Concerning Human Understanding*. Oxford: Oxford University Press.
- Lockie, Robert. (2014). Three Recent Frankfurt Cases. *Philosophia*, 42(4), 1005-1032.
- McKenna, Michael. (2003). Robustness, Control, and the Demand for Morally Significant Alternatives: Frankfurt Examples with Oodles and Oodles of Alternatives. In D. Widerker & M. McKenna (Eds.), *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities* (pp. 201-217). Aldershot: Ashgate.
- McKenna, Michael. (2005). Where Frankfurt and Strawson Meet. *Midwest Studies in Philosophy*, 29(1), 163-180.
- McKenna, Michael. (2008). Frankfurt's Argument Against Alternative Possibilities: Looking Beyond the Examples. *Noûs*, 42(4), 770-793.
- Mele, Alfred R. (2006). *Free Will and Luck*. Oxford: Oxford University Press.
- Mele, Alfred R., & Moser, Paul K. (1994). Intentional action. *Noûs*, 28(1), 39-68.
- Mele, Alfred R., & Robb, David. (1998). Rescuing Frankfurt-style Cases. *Philosophical Review*, 107(1), 97-112.
- Mele, Alfred R., & Robb, David. (2003). Bbs, Magnets and Seesaws: The Metaphysics of Frankfurt-style Cases. In D. Widerker & M. McKenna (Eds.), *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities* (pp. 127-138). Aldershot: Ashgate.
- Mele, Alfred R., & Sverdlik, Steven. (1996). Intention, Intentional Action, and Moral Responsibility. *Philosophical Studies*, 82(3), 265-287.
- Nelkin, Dana. (2008). Responsibility and Rational Abilities: Defending an Asymmetrical View. *Pacific Philosophical Quarterly*, 89(4), 497-515.
- Nelkin, Dana. (2011). *Making Sense of Freedom and Responsibility*. New York: Oxford University Press.
- O'Connor, Timothy. (1996). Why Agent Causation? *Philosophical Topics*, 24, 143-158.
- O'Connor, Timothy. (2000). *Persons and Causes: the Metaphysics of Free Will*. New York: Oxford University Press.



- Oshana, Marina. (1997). Ascriptions of Responsibility. *American Philosophical Quarterly*, 34(1), 71-83.
- Palmer, David. (2011). Pereboom on the Frankfurt Cases. *Philosophical Studies*, 153(2), 261-272.
- Palmer, David. (2013a). Capes on the W-Defense. *Philosophia*, 41(2), 555-566.
- Palmer, David. (2013b). The Timing Objection to the Frankfurt Cases. *Erkenntnis*, 78(5), 1011-1023.
- Pereboom, Derk. (1995). Determinism Al Dente. *Noûs*, 29(1), 21-45.
- Pereboom, Derk. (2000). Alternative Possibilities and Causal Histories. *Philosophical Perspectives*, 34(14), 119-138.
- Pereboom, Derk. (2001). *Living Without Free Will*. Cambridge: Cambridge University Press.
- Pereboom, Derk. (2005). Defending Hard Incompatibilism. *Midwest Studies in Philosophy*, 29(1), 228-247.
- Pereboom, Derk. (2008). Defending Hard Incompatibilism Again. In N. Trakakis & D. Cohen (Eds.), *Essays on Free Will and Moral Responsibility* (pp. 1-33). Cambridge: Cambridge Scholars Press.
- Pereboom, Derk. (2009). Further Thoughts About a Frankfurt-Style Argument. *Philosophical Explorations*, 12(2), 109 – 118.
- Pereboom, Derk. (2012). Frankfurt Examples, Derivative Responsibility, and the Timing Objection. *Philosophical Issues*, 22(1), 298-315.
- Pereboom, Derk. (2013). Optimistic Skepticism about Free Will. In P. Russell & O. Deery (Eds.), *The Philosophy of Free Will: Selected Contemporary Readings*. New York: Oxford University Press.
- Pereboom, Derk. (2014). *Free Will, Agency, and Meaning in Life*. New York: Oxford University Press.
- Robinson, Michael. (2012). Modified Frankfurt-Type Counterexamples and Flickers of Freedom. *Philosophical Studies*, 157(2), 177-194.
- Robinson, Michael. (2014). The Limits of Limited-Blockage Frankfurt-Style Cases. *Philosophical Studies*, 169(3), 1-18.
- Scanlon, Thomas. (1998). *What We Owe to Each Other*. Cambridge: Harvard University Press.
- Sinnott-Armstrong, Walter. (1988). *Moral Dilemmas*. Oxford: Blackwell Publishers.

- Smith, Angela M. (2005). Responsibility for Attitudes: Activity and Passivity in Mental Life. *Ethics*, 115(2), 236-271.
- Smith, Michael. (2003). Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion. In S. Stroud & C. Tappolet (Eds.), *Weakness of Will and Practical Irrationality* (pp. 17-38). Oxford: Clarendon Press.
- Speak, Daniel. (2005). PAPistry: Another Defense. *Midwest Studies in Philosophy*, 29(1), 262-268.
- Strawson, P. F. (1962). Freedom and Resentment. In *Freedom and Resentment and Other Essays* (pp. 1-28). New York: Routledge.
- Stump, Eleonore. (1996). Libertarian Freedom and the Principle of Alternative Possibilities. In J. Jordan & D. Howard-Snyder (Eds.), *Faith, Freedom, and Rationality: Philosophy of Religion Today* (pp. 73-88). Lanham: Rowman & Littlefield.
- Stump, Eleonore. (1999). Dust, Determinism, and Frankfurt. *Faith and Philosophy*, 16(3), 413-422.
- Stump, Eleonore. (2003). Moral Responsibility Without Alternative Possibilities. In D. Widerker & M. McKenna (Eds.), *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities* (pp. 139-158). Aldershot: Ashgate.
- Timpe, Kevin. (2008). *Free Will: Sourcehood and Its Alternatives*. London: Continuum.
- van Inwagen, Peter. (1983). *Essay on Free Will*. Oxford: Clarendon Press.
- van Inwagen, Peter. (1999). Moral Responsibility, Determinism, and the Ability To Do Otherwise. *Journal of Ethics*, 3(4), 343-351.
- van Inwagen, Peter. (2000). Free Will Remains a Mystery. *Philosophical Perspectives*, 14, 1-20.
- van Inwagen, Peter. (2008). How to Think About the Problem of Free Will. *Journal of Ethics*, 12(3/4), 327-341.
- van Someren Greve, Rob. (2014). 'Ought', 'Can', and Fairness. *Ethical Theory and Moral Practice*, 17(5), 913-922.
- Vihvelin, Kadri. (2004). Free Will Demystified: A Dispositional Account. *Philosophical Topics*, 32(1/2), 427-450.
- Watson, Gary (Ed.). (2004). *Agency and Answerability: Selected Essays*. New York: Oxford University Press.
- Widerker, David. (1991). Frankfurt on 'Ought Implies Can' and Alternative Possibilities. *Analysis*, 51(4), 222 - 224.

- Widerker, David. (1995). Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities. *The Philosophical Review*, 104(2), 247-261.
- Widerker, David. (2000). Frankfurt's Attack on the Principle of Alternative Possibilities: A Further Look. *Philosophical Perspectives*, 14(s14), 181-202.
- Widerker, David. (2003). Blameworthiness and Frankfurt's Argument Against the Principle of Alternative Possibilities. In D. Widerker & M. McKenna (Eds.), *Freedom, Responsibility, and Agency: Essays on the Importance of Alternative Possibilities* (pp. 53-74). Burlington: Ashgate.
- Widerker, David. (2005). Blameworthiness, Non-robust Alternatives, and the Principle of Alternative Expectations. *Midwest Studies in Philosophy*, 29(1), 292-306.
- Widerker, David. (2006). Libertarianism and the Philosophical Significance of Frankfurt Scenarios. *The Journal of Philosophy*, 103(4), 163-187.
- Widerker, David. (2009). A Defense of Frankfurt-friendly Libertarianism. *Philosophical Explorations*, 12(2), 87-108.
- Widerker, David, & Goetz, Stewart. (2013). Fischer Against the Dilemma Defense: The Defense Prevails. *Analysis*, 73(2), 283-295.
- Widerker, David, & McKenna, Michael (Eds.). (2003). *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*. Aldershot: Ashgate.
- Wolf, Susan. (1990). *Freedom Within Reason*. New York: Oxford University Press.
- Yaffe, Gideon. (1999). 'Ought' implies 'Can' and the Principle of Alternate Possibilities. *Analysis*, 59(263), 218-222.
- Zagzebski, Linda. (2000). Does Libertarian Freedom Require Alternate Possibilities? *Noûs*, 34(s14), 231-248.
- Zimmerman, Dean. (2007). *Oxford Studies in Metaphysics*. New York: Oxford University Press.
- Zimmerman, Michael J. (1988). *An Essay on Moral Responsibility*. Totowa, NJ: Rowman & Littlefield.
- Zimmerman, Michael J. (1997). A Plea for Accuses. *American Philosophical Quarterly*, 34(2), 229-243.
- Zimmerman, Michael J. (2003). The Moral Significance of Alternate Possibilities. In D. Widerker & M. McKenna (Eds.), *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities* (pp. 301-325). Aldershot: Ashgate.