2013

Moral Responsibility and Its Alternatives

Michael Robinson
MORAL RESPONSIBILITY AND ITS ALTERNATIVES

By

MICHAEL ROBINSON

A Dissertation submitted to the Department of Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Degree Awarded: Summer Semester, 2013

Copyright © 2013 Michael Robinson All Rights Reserved
Michael Robinson defended this dissertation on June 25, 2013.
The members of the supervisory committee were:

Alfred Mele
Professor Directing Dissertation

Aline Kalbian
University Representative

Randolph Clarke
Committee Member

J. Piers Rawling
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the dissertation has been approved in accordance with university requirements.
For my parents, Earl and Brenda
ACKNOWLEDGMENTS

I am grateful to Al Mele, Randy Clarke, Piers Rawling, and Aline Kalbian for serving on my dissertation committee. I would like to thank Randy Clarke, Michael McKenna, and John Fischer for their many helpful comments, both in writing and in person, on large portions of this dissertation. I am deeply grateful for their encouragement, as well as for that of David McNaughton, whose friendship has been a constant source of support. I would also like to thank Joe Long, Travis Rodgers, and Justin Capes for the many hours spent discussing the ideas and arguments in this dissertation while playing disc golf on Saturday mornings in Tom Brown Park. My greatest debt in this dissertation is to my advisor, Al Mele, who seemed never to tire of providing helpful and extensive feedback on earlier drafts of each chapter, from which I benefitted immensely. I know no better philosopher, nor do I know of one who is more generous with his time. I consider myself incredibly fortunate to have worked closely with him these past years. I am a much better philosopher and writer because of it, and I will always be grateful for the experience. Last, but not least, I would like to thank my parents, Earl and Brenda, for a lifetime of loving encouragement and support and for always viewing my education as a sound investment.
TABLE OF CONTENTS

Abstract ......................................................................................................................................... vii

1. INTRODUCTION ................................................................................................................1

2. MODIFIED FRANKFURT-TYPE COUNTEREXAMPLES AND FLICKERS OF
   FREEDOM ............................................................................................................................10
   2.1 Introduction ................................................................................................................10
   2.2 Traditional Frankfurt-Style Counterexamples ..........................................................11
   2.3 Getting Rid of Prior Signs .......................................................................................12
   2.4 For What is Jones Morally Responsible? ...............................................................14
   2.5 An Artificial Separation? .......................................................................................18
   2.6 A More Modest Approach ....................................................................................22
   2.7 The Robustness Objection .....................................................................................25
   2.8 Conclusion ...............................................................................................................28

3. THE LIMITS OF LIMITED-BLOCKAGE FRANKFURT-STYLE CASES ......................31
   3.1 Introduction ...............................................................................................................31
   3.2 The Trouble with Traditional Frankfurt-Style Cases .............................................32
   3.3 The Limited-Blockage Strategy .............................................................................34
   3.4 Two Kinds of Frankfurt-Style Attack ....................................................................38
   3.5 Step One: The Connection between PAP and PSA .............................................40
   3.6 Step Two: PSA* and Brain Malfunction ..............................................................44
   3.7 Conclusion ...............................................................................................................50

4. ROBUST FLICKERS OF FREEDOM ............................................................................52
   4.1 Introduction ...............................................................................................................52
   4.2 Frankfurt-Style Cases and Flickers of Freedom ....................................................54
   4.3 Why Must Flickers Be Robust? ...............................................................................60
   4.4 What Does Robustness Require? ..........................................................................61
   4.5 Robust Flickers of Freedom ....................................................................................63
   4.6 An Objection and a Reply ......................................................................................69
   4.7 Conclusion ...............................................................................................................73

5. ALTERNATIVE CONCEPTIONS OF MORAL RESPONSIBILITY ................................74
   5.1 Introduction ...............................................................................................................74
   5.2 Deflationary Accounts of Moral Responsibility ....................................................75
   5.2.1 Smart on Responsibility ....................................................................................76
   5.2.2 Scanlon on Responsibility ................................................................................77
   5.2.3 Smith on Responsibility ....................................................................................82
   5.3 Inflationary Accounts of Moral Responsibility ....................................................85
   5.3.1 Heaven-and-Hell Responsibility ....................................................................86
5.3.2 Clarke’s Objection to Strawson .................................................................89
5.3.3 The Real Problem with Heaven-and-Hell Responsibility ..................92
5.3.4 Why This Matters ..............................................................................96
5.4 Conclusion .........................................................................................99

6. CONCLUSION ......................................................................................101

REFERENCES ............................................................................................102

BIOGRAPHICAL SKETCH ............................................................................107
It has long been held that a person is morally responsible for what she has done only if she could have done otherwise. This is commonly known as the Principle of Alternative Possibilities (PAP). In this dissertation I defend PAP against two main lines of attack. The first comes from a class of putative counterexamples to PAP devised by Harry Frankfurt, commonly known as Frankfurt-style cases. The second line of attack I consider comes from various attempts in recent years to reconceive the nature of moral responsibility in a way that straightforwardly entails that alternative possibilities are not required for moral responsibility. I argue that these challenges are unsuccessful in undermining the view that moral responsibility requires the ability to do otherwise and, in doing so, attempt to clarify the moral significance of possessing alternatives of this sort.
CHAPTER ONE

INTRODUCTION

It has long been held that a person is morally responsible for what he has done only if he could have done otherwise. This is commonly known as the Principle of Alternative Possibilities (PAP). The thought that being morally responsible for something we have done requires that it was possible for us not to have done it is deeply intuitive. Indeed, it seems to be something of a commonplace. If I break my promise to pick you up from the airport, you will likely be predisposed to blame me for this, unless I have an adequate excuse of some sort. One kind of excuse I might have would involve demonstrating that I had a justificatory reason for breaking my promise. Perhaps a neighborhood child had been seriously injured and I was the only one around who could take him to the hospital, or perhaps my boss threatened to fire me if I did not stay late at work to help her finish some last minute project. Excuses of this sort would generally show that it would be inappropriate for you to continue to blame me for breaking my promise to you—not because I was not morally responsible for my actions but, rather, because I was morally justified in what I did. A different kind of excuse I might have would involve demonstrating that I had an exculpatory reason for failing to keep my promise. Perhaps it was I who (through no fault of my own) had been seriously injured and taken (say, unconscious or lame) to the hospital, or perhaps burglars had chained me up in the basement just prior to the time I would have needed to leave to pick you up from the airport. Excuses of this sort would also generally show that it would be inappropriate for you to continue to blame me for failing to keep my promise to you. Importantly, however, they do not do this by showing that omitting to pick you up at the airport was the right thing to do; rather, they do so by showing that I was not at all morally responsible for what I did. And, on the face of it at least, the reason that the latter kind of excuse does this while the former kind of excuse does not seems plain enough: Unlike the former excuses, the latter excuses show that it was not possible for me to have kept my promise. Thus, the idea that a person’s having been unable to avoid doing something she did (or
having been unable to do something she failed to do) is sufficient to preclude her from being morally responsible for this appears to be firmly embedded in our social practices.

Beyond its apparent place in the everyday interactions of ordinary people, PAP has played a central role in the longstanding philosophical debate concerning the relationship between causal determinism and moral responsibility. Causal determinism is the view that all events that occur are causally determined (by prior events together with the laws of nature) to occur in such a way that a complete description of the state of the world at a given time and a complete description of the natural laws jointly imply all other true descriptions of the world at all other times. Given the past and the laws, there is only one possible future. Whereas compatibilists maintain that agents can be morally responsible for (at least some of) their actions even if they are causally determined, incompatibilists insist that the truth of causal determinism would preclude agents from being morally responsible for anything they do. One of the most common ways that incompatibilists have sought to support their position is by appealing to PAP (and other similar principles). The reason that causal determinism is incompatible with moral responsibility, they have argued, is precisely because of the effect that determinism would have on agents’ alternative possibilities: it would eliminate them. If causal determinism is true, then it is never the case that agents could have done otherwise than perform the actions they perform—which, according to PAP, entails that agents are never morally responsible for anything they do. Interestingly, for much of the history of this debate, the compatibilist response to this argument for incompatibilism was to reject the claim that determinism eliminated alternative possibilities.\(^1\) Thus, though there was disagreement over whether an agent’s being causally determined to do what he did implies that he could not have done otherwise, that an agent’s having been able to do otherwise was required in order for him to be morally responsible for what he did was not in dispute. PAP was widely viewed by both parties as simply too commonsensical to be rejected.

In light of its strong intuitive appeal, as well as its apparent deep-rootedness in our social practices, it is perhaps surprising that this principle would stand in much need of defense. Over

---

1 In the current debate, compatibilists who accept that moral responsibility for one’s actions requires that one could have done otherwise are commonly referred to as “classical compatibilists.” Classical compatibilists defend a conditional analysis of ‘could have done otherwise claims’, according to which to say that an agent could have done otherwise than perform some action he performed just means that the agent would have done otherwise if he had wanted (or chosen) to do otherwise. See, for example, Nielsen 1971 and Ayer 1954. For criticisms of this view, see Chisholm 1964, Austin 1961, and Lehrer 1968.
the past several decades, however, many philosophers (perhaps a majority at this point) have become convinced that PAP ought to be rejected. The aim of this dissertation is to defend PAP from two main lines of attack. The first comes from a class of putative counterexamples developed by Harry Frankfurt, commonly referred to as Frankfurt-style cases. The second challenge I consider comes from various attempts in recent years to reconceive the nature of moral responsibility in a way that settles (in the negative) the question of the necessity of alternative possibilities for moral responsibility.

No one has been more influential in the recent move away from PAP than Harry Frankfurt (1969). In his now famous paper, “Alternate Possibilities and Moral Responsibility,” Frankfurt argued that the ability to do otherwise is not in fact a necessary condition for moral responsibility, for we can construct scenarios in which people appear to be morally responsible for performing certain actions even though it was not possible for them to have done otherwise:

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 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. (Frankfurt 1969, pp. 835-836)

Frankfurt argues that cases of this kind provide us with good grounds for rejecting PAP. Here, for instance, it is not possible for Jones to bring it about that he does not perform the action he performs. Thus, if PAP were correct in claiming that moral responsibility requires the ability to do otherwise, then Jones would not be morally responsible for what he does. But since Jones acts on his own and Black plays no causal role in Jones’s decision, it seems that there is no good
reason to deny that Jones is genuinely morally responsible for what he does. Therefore, we ought to reject PAP.

Cases of this sort have proved effective in persuading many philosophers to abandon the ability to do otherwise as a necessary condition for moral responsibility. To neutralize the threat these cases pose to PAP, defenders of this principle will need to provide good reasons for overcoming the apparent intuition that agents in these cases are morally responsible for what they do or else show how PAP can accommodate this claim. In Chapters Two, Three, and Four of this dissertation I adopt the latter tack.

A great deal of attention has been paid recently to the claim that Frankfurt-type counterexamples are guilty of begging the question against incompatibilists. In particular, it is argued that traditional Frankfurt cases—which fundamentally involve the presence of a counterfactual intervener to guarantee that an agent will act in a certain way—depend for their success on there being a perfectly reliable indicator (or prior sign) of what an agent will freely do if left to act on his own. Such indicators, however, presuppose a deterministic relation between an agent’s free action and its causal antecedents—a relation incompatibilists flatly deny. Objections of this sort have given rise to considerable efforts to construct alternative Frankfurt-type counterexamples that do not rely on prior signs and so do not presuppose determinism in a way that incompatibilists would find objectionable.

In Chapter Two I develop two lines of argument intended to show that the principle of alternative possibilities is not falsified by Frankfurt-style cases. The first consists in arguing that the actions that agents could not have avoided performing in newly modified Frankfurt cases are not the same actions for which they are morally responsible. The second takes the principle of alternative possibilities to be a thesis about things for which one is basically (or non-derivatively) morally responsible, which, it is argued, are things that one could have avoided doing in alternative Frankfurt scenarios. Insofar as either of these attempts is successful, the upshot is that, even if these new cases successfully manage to sidestep worries about the question-begging nature of prior signs, they fail to constitute genuine counterexamples to the principle of alternative possibilities.

Both of the arguments I offer in Chapter Two are versions of the so-called “flicker strategy.” Defenses of this sort emphasize the fact that Frankfurt-style scenarios, though they succeed in ruling out some of the options open to agents in ordinary situations, do not eliminate
all alternative possibilities.\textsuperscript{2} Returning to Frankfurt’s original scenario, although there is nothing Jones can do to prevent himself from performing the action he in fact performs (the one Black wants him to perform), it remains open whether or not he will do this \textit{on his own} (a point Frankfurt himself readily acknowledges), and in these alternatives the flickers of freedom can be found. Thus, PAP can accommodate the claim that agents in these cases are morally responsible for what they do.

The flicker strategy quickly fell out of favor, owing in large part to an objection by John Martin Fischer. Fischer argued that even if it remains open to agents in Frankfurt-type counterexamples not to perform a certain action, A, on their own, this is no help to defenders of PAP because “this alternative possibility is not sufficiently \textit{robust} to ground the relevant attributions of moral responsibility” (1994, p. 140). In order for the existence of alternative possibilities in Frankfurt-type counterexamples to neutralize the threat to PAP, the alternatives must be able to account for why an agent would, merely in virtue of possessing these alternatives, be thereby morally responsible for what he does. But why is it that the possibility of not A-ing on one’s own in Frankfurt-type counterexamples is not sufficiently robust to ground the relevant attributions of moral responsibility? The reason for this, according to Fischer, is that the addition of alternatives of this sort in no way adds to the control that the agent has over what he does. Although it remains open to agents in these cases not to A on their own but, rather, to A as a result of outside intervention, whether or not they do so is not something over which they have control. It is possible for agents in these cases to display some sort of sign, indicating that they will not freely act in the desired way on their own, but, by hypothesis, the nature of these signs is such that they are involuntary, and so not within the agents’ control. If it were open to these agents to indicate that they will not freely A on their own by forming an intention of some sort (say, an intention not to A), then it would be plausible to suppose that these alternatives could be robust in the way Fischer requires. Since, however, the indication that they will not A on their own is more like an involuntary twitching than a voluntary act of decision-making, the fact that it remains open to agents in Frankfurt cases to display a sign of this sort is unable to do the work set out for it by flickers-of-freedom strategists.

At the end of Chapter Two I argue that part of the significance of the rejection of prior signs in the construction of Frankfurt-type counterexamples for the flickers of freedom defense

\textsuperscript{2} Early versions of the flicker strategy can be found in van Inwagen 1978 and Naylor 1984.
of PAP is that it ensures that up until the moment of decision it is possible for agents in those cases to realize one of the alternative possibilities that remain even in modified Frankfurt cases. Because these possibilities are available at the moment of choice (the locus of free action), and so within the range of one’s voluntary control, it is no longer obvious that they are insufficiently robust to help ground ascriptions of moral responsibility—at least, not for the reasons Fischer provides. Thus, those wishing to show that the alternatives that remain open to agents in Frankfurt-style cases are not sufficiently robust will need to find new ways to support this claim.

In Chapter Three I focus on a recent class of “limited-blockage” Frankfurt-style cases developed by Michael McKenna (2003) that purport to answer each of these challenges. Philosophers employing Frankfurt-style cases to challenge PAP have mostly sought to construct scenarios that eliminate as many of an agent’s alternatives as possible—and all alternatives at the moment of action, within the agent’s control—but without causally determining the agent’s actions (lest this call into question the agent’s moral responsibility). One of the chief difficulties for this traditional approach is that the closer one gets to eliminating absolutely all alternative possibilities the more it appears that agents’ actions in these cases are causally determined. Limited-blockage versions of these cases are intended to sidestep this worry by blocking all and only those alternatives that are intrinsically relevant to moral responsibility (that is, robust alternatives) while leaving open all other alternatives, including a significant range of alternatives that are within the agent’s voluntary control at the moment of action. I argue that, owing to the fact that omissions (and not just actions) are capable of constituting robust alternative possibilities, limited-blockage cases cannot avoid collapsing into the more traditional sort of Frankfurt-style case to which they are meant to be an alternative and so are vulnerable to the very same concerns they are meant to avoid.

Chapter Four draws on two key lessons from Chapters Two and Three—that agents in these cases are basically morally responsible deciding on their own to act as they do and that nonintentional omissions can be within agents’ direct voluntary control—to develop a general response to the robustness challenge, one that can be applied to any present or future Frankfurt-style case. As we have seen, critics of the flicker strategy have argued that the alternative possibilities that remain open to agents in Frankfurt-style cases—including the ability to omit to perform certain actions on their own—are not adequate to the task of rebuffing Frankfurt-style attack. In order to neutralize the threat facing PAP, it is not enough, they maintain, for agents in
these cases to have alternative possibilities—the alternatives must be robust. That is, they must be morally significant in a way that helps to account for the agents’ moral responsibility for what they actually do. Insofar, then, as Frankfurt-style cases are able to restrict agents’ ability to do otherwise to alternatives that are insufficiently robust in this sense, they will be successful in undermining PAP (assuming these agents are morally responsible for what they do). Eliminating all alternative possibilities, including those that are not robust, they argue, is not required.

My aim in this chapter is to consider whether there is any good reason for thinking both (i) that the success of the flicker strategy requires appealing to distinctly robust alternative possibilities and (ii) that there are (or at least can be) Frankfurt-style cases that eliminate all such alternatives (again, without undermining agents’ moral responsibility for what they do). I argue that whereas demonstrating that Frankfurt-style cases fail to constitute genuine counterexamples to PAP does not require showing that at least some of the alternatives open to agents in these cases are distinctly robust, there is another kind of Frankfurt-style attack where the demand for robustness is appropriate. There the goal in offering Frankfurt-style cases is not to falsify PAP (perhaps because this is viewed as a less promising venture) but, rather, to undercut its motivation by challenging certain intuitions taken to be the grounds for this principle. As I shall further argue, however, even if all that is needed to demotivate (though not falsify) PAP in this way are Frankfurt-style scenarios that manage to eliminate all robust alternative possibilities, there are no (nor can there be) cases of this sort. So long as it remains up to agents in these cases whether to on their own do the things for which they are morally responsible, they will continue to possess robust alternative possibilities. And since having it be up to the agents whether they perform the relevant actions on their own is a crucial feature of Frankfurt-style cases—one which cannot be eliminated without undermining the judgment that they are genuinely morally responsible for what they do—the demand for robust flickers of freedom is one that this particular version of the flicker strategy will always be able to meet.

Chapter Five focuses on a challenge to PAP posed by recent attempts to reconceive the nature of moral responsibility in a way that straightforwardly implies that whether an agent is morally responsible for something he has done has nothing to do with whether he had alternative possibilities. Consider, for instance, the conception of moral evaluation offered by J. J. C. Smart (1961). According to Smart, in general, to praise or dispraise a thing is just to grade it, that is, to evaluate it according to some standard. A thing is praised when it receives a positive evaluation
and dispraised when it receives a negative evaluation, just as a student might be praised for her mathematical skills and an athlete might be dispraised for his incompetence on the field; and the same is true of moral qualities and actions. He writes:

Just as we may praise or dispraise a woman for her figure, a footballer for his fleetness or slowness of foot, a lecturer in philosophy for his intelligence or lack of intelligence, and a writer for clarity or obscurity, so naturally enough, we may praise or dispraise a man for his honesty or dishonesty, truthfulness or untruthfulness, kindness or unkindness, and so on. In this sense of ‘praise’ we may praise moral qualities and moral actions in exactly the same way as we may praise beauty, intelligence, agility or strength. (304)

Of course, although we may dispraise a person for his ugliness, we do not typically blame him for it. The reason for this, Smart maintains, is that, in addition to grading, to blame a person for something implies that it is both open to modification and something for which she is responsible (in the sense that it is the result of certain choices she made, and not in any deeply metaphysical sense). Aside from this, however, blame should be viewed in just the same way as dispraise. “Blame in this sense,” he says, “can be just as dispassionate as dispraise of a woman’s nose: it is just a grading plus an ascription of responsibility” (305).

Plainly, alternative possibilities are not needed in order to be morally responsible according to Smart’s conception of responsibility. Thus, if Smart is right about the nature of moral responsibility, then the question whether it requires the ability to do otherwise simply fades away. More than that, the entire debate surrounding possibilities of this sort seems not to make any sense. In this chapter I seek to draw attention to two general ways that some have attempted to revise the traditional understanding of what it is to be morally responsible, each of which implies that the ability to do otherwise is a non-issue. The first (and more common) way is to defend a *deflationary account* of moral responsibility, according to which responsibility does *not* entail that one is (or even can be) deserving of any amount of blame or praise, punishment or reward for things one has done. The second way is to propose an *inflationary account* of moral responsibility, according to which responsibility entails not only that one is deserving of some amount of praise or blame, punishment or reward but that one is (or at least could be) deserving of an *infinite* amount of punishment or reward.
One problem that is shared by each of these conceptions of moral responsibility, I argue, is that it is unclear why anyone would think that bearing these kinds of responsibility has anything to do with the ability to do otherwise. Insofar as a given conception is unable either to give an adequate account of the force and fairness typically associated with judgments of moral responsibility or to allow for the intelligibility of the longstanding debate over the necessity of alternative possibilities, so much the worse for that conception of responsibility.

At any rate, it is important to identify those places in the debate where arguments rely on a particular notion of moral responsibility, especially when that notion is not shared by one’s interlocutors. When this does not happen, the remaining ambiguity tends to result either in both sides simply talking past each other or in one side trading on the ambiguity in ways that are illicit. The matter of the relationship between moral responsibility and the ability to do otherwise is difficult but important; it should not be decided with one side simply moving the goal posts and declaring victory.
2.1 Introduction

A great deal of attention has been paid recently to the claim that Frankfurt-type counterexamples, whose aim is to falsify the view that moral responsibility requires alternative possibilities, are guilty of begging the question against incompatibilists. In particular, it is argued that traditional Frankfurt cases, which fundamentally involve the presence of a counterfactual intervener to guarantee that an agent will act in a certain way, depend for their success on there being a perfectly reliable indicator (or prior sign) of what an agent will freely do if left to act on his own. Such indicators, however, seem to presuppose a deterministic relation between an agent’s free action and its causal antecedents—a relation incompatibilists flatly reject. Objections of this sort have given rise to considerable efforts to construct alternative Frankfurt-type counterexamples that do not rely on prior signs of this kind and so do not presuppose determinism in a way that incompatibilists should find objectionable.

One consequence of this shift in the way Frankfurt-type counterexamples are formulated is that it provides an opportunity for the forceful resurgence of certain versions of the so-called flicker defense of alternative possibilities. Defenses of this sort proceed by arguing that even if Frankfurt cases succeed in eliminating many of the alternatives generally available to agents in ordinary circumstances, some alternatives (the flickers of freedom) remain—alternatives which

* From *Philosophical Studies* (2012) 157, 177-194. Reprinted by permission of the publisher, for which I am grateful. For helpful comments on earlier versions of this chapter, I wish to thank Al Mele, Michael McKenna, Randy Clarke, Walt Schaller, Sara Chant, Travis Rodgers, and an anonymous referee for *Philosophical Studies*. A much earlier version of the first half of this chapter was presented at the 2007 Pacific-Mountain Regional Meeting of the Society of Christian Philosophers. I am grateful to the audience there, and especially to Dan Speak, for their feedback.
are necessary for moral responsibility.\textsuperscript{3} As a response to traditional Frankfurt scenarios, the flicker defense has been criticized on the grounds that the alternatives that persist in cases of this sort, since they are not things over which an agent has any considerable control, cannot plausibly be held to contribute to an agent’s moral responsibility and so are unable to blunt Frankfurt-style attack. Whatever the merits of this objection in the context of traditional Frankfurt scenarios, when employed in defense of claims based on modified Frankfurt cases, where the removal of prior signs has the effect of returning these alternatives to within the scope of an agent’s direct voluntary control, it has no force.

In this chapter I develop two versions of the flicker defense of alternative possibilities, indicate their advantages over other versions of this strategy, and defend them against objections. Insofar as either of these is successful, the upshot is that even if some modified Frankfurt-type counterexamples effectively manage to sidestep worries concerning the question-begging nature of prior signs, they still fail to falsify the view that moral responsibility requires alternative possibilities.

\textbf{2.2 Traditional Frankfurt-Type Counterexamples}

It has long been held that a person is morally responsible for what he has done only if he could have done otherwise. This is commonly known as the Principle of Alternative Possibilities (PAP). Harry Frankfurt (1969) has famously argued that the ability to do otherwise is not in fact a necessary condition for moral responsibility; for we can think of cases in which people appear to be morally responsible for things they have done even though it was not possible for them to have done otherwise.

Consider the following Frankfurt-type counterexample. Jones is a disgruntled postal worker who goes to work on Monday morning unaware that his boss intends to fire him just before noon. Black, an evil neuroscientist, outraged by the increased cost of domestic postage and aware of what Jones’s boss plans to do, hopes that getting fired will make Jones so angry that he decides to burn the place to the ground. Black would prefer that Jones do this on his own. Not wanting to take any chances, however, Black has, unbeknownst to Jones, implanted a special device in Jones’s brain which will enable him to cause Jones to decide to set fire to the post

\textsuperscript{3} The term ‘flickers of freedom’ was coined by John Martin Fischer (1994, pp. 137-47) to refer to those alternatives that remain open to agents in Frankfurt-type counterexamples.
office (and do so) in the event that Jones gives any indication he will not do so on his own. If just before noon Black (somehow) is able to tell that Jones is not going to set fire to the post office on his own, Black will activate the device via remote control, sending electrical impulses to specified parts of Jones’s brain, causing him to decide to set fire to the post office and then do so. As it happens, upon being fired, and without any interference from Black, Jones flips out, punches his boss in the face, and burns the place to the ground. In short, Jones goes postal. Let us call this case “Post Office.”

Frankfurt argues that cases of this kind provide us with good grounds for rejecting PAP. In Post Office, for example, it is not possible for Jones to avoid deciding to set fire to the post office. Thus, if PAP were correct in claiming that moral responsibility requires that one could have done otherwise, then Jones would not be morally responsible for what he does. Since Jones acts on his own, however, and Black plays no causal role in Jones’s decision to set the post office on fire, it seems that Jones is genuinely morally responsible for what he does. Therefore, we ought to reject PAP.

2.3 Getting Rid of Prior Signs

Traditional Frankfurt-type counterexamples like Post Office have come under heavy attack in recent years on the grounds that they are guilty of begging the question against incompatibilists (see Widerker 1995, Kane 1996, Ginet 1996). The problem has to do with the way in which the counterfactual intervener is able to know what the agent will freely do if left to act on his own. In Post Office, whether Black intervenes depends on whether Jones shows some indication that he will not freely decide on his own to set fire to the post office. But what sort of indication might this be? Suppose Black knows that if Jones blushes at some point prior to noon, then at noon Jones will freely decide to set the place ablaze. In this case, Jones’s blushing will serve as a prior sign to Black that Jones will do what Black wants him to do without any interference being necessary.

The difficulty with this sort of approach, it is argued, comes in grounding the truth of the claim that if Jones blushes he will freely decide to set fire to the post office. If Jones’s blushing makes it merely probable (say, given his past behavior) that he will set the blaze himself, then up until noon it remains causally open to him not to decide to burn down the post office after all. Insofar as this is the case, however, Jones continues to possess alternative possibilities, and so his
bearing moral responsibility for his actions poses no threat to PAP. On the other hand, if Jones’s blushing at some point prior to noon will itself deterministically lead to his deciding to set the fire, or if it is indicative of some other such event, then it will no longer remain causally open to Jones not to decide to set the post office on fire. But since having one’s actions deterministically caused precludes one from being morally responsible for them, according to incompatibilists, that Jones is nevertheless morally responsible for what he does cannot be asserted without begging any questions. This, then, is the dilemma facing traditional Frankfurt-type counterexamples: Either Jones’s blushing (or that which it indicates) does not deterministically lead to his deciding to set fire to the post office, in which case he retains alternative possibilities regarding his actions, or his so deciding is in fact causally determined by prior events of which his blushing is a sign, in which case the example loses all force for incompatibilists, who should simply reject the claim that Jones is morally responsible for what he does. Either way, PAP is safe.

This particular line of objection has prompted a number of attempts to modify traditional Frankfurt cases in a way that enables them to do the work set out for them by Frankfurt without employing the use of prior signs. One such attempt, offered by Alfred Mele and David Robb (1998, 2003), suggests that traditional Frankfurt-type scenarios, such as Post Office, be modified in the following way. So as not to beg any questions against incompatibilists, let us suppose that Jones’s brain operates in an indeterministic manner. Further suppose that instead of waiting to see whether Jones gives some indication he will set fire to the post office on his own, Black implants a deterministic device in Jones’s brain that at noon will issue in a decision to set the place ablaze unless Jones decides on his own to set the fire. In the event that, at noon, Jones’s own cognitive faculties issue in a decision to set fire to the post office, the deterministic

---

4 It should be noted that, relying as it does on the decidedly incompatibilist intuition that determinism and moral responsibility are mutually exclusive, this is not a line of argument available to compatibilist defenders of PAP. Compatibilists who wish to resist the conclusion of Frankfurt’s argument cannot do so on the grounds that it is question-begging to assert that agents can be morally responsible for their actions despite their being causally determined.

5 For arguments to the effect that there is in fact nothing illicit in the way traditional Frankfurt-type counterexamples make use of prior signs, see Fischer 1999 and Haji and McKenna 2004.

6 See, for example, Haji 1998; Hunt 2000; McKenna 2003; Mele and Robb 1998, 2003; Pereboom 2000; and Stump 1996.
mechanism implanted by Black will at that very moment be causally preempted from issuing in a decision. The significance of the preemption’s occurring precisely at noon is that it guarantees that at no point leading up to and including noon is it possible for Jones to avoid deciding to set the fire. Furthermore, since this deterministic mechanism is screened off from Jones’s consciousness and plays no causal role in his deciding to act as he does, he looks to be genuinely morally responsible for his actions. Call this “Mod ified Post Office.”

Other attempts to modify traditional Frankfurt scenarios employ a variety of means to ensure that agents in such cases cannot avoid doing what they do without presupposing determinism. None of what follows, however, turns on any of the peculiarities of Mele and Robb’s particular approach. Any conclusions drawn about Modified Post Office should therefore be seen as applying equally to all modified versions of Frankfurt-type counterexamples that are able to avoid the dilemma laid out above.

It is perhaps worth noting that Mele and Robb’s particular type of modified Frankfurt scenario, though it does not rely on the use of prior signs, is itself the subject of some controversy. The chief complaint about this specific sort of example has to do with worries surrounding the theoretical possibility of the kind of \textit{occurrent} causal preemption (or “trumping”) to which it appeals.\footnote{For their reply to this and other concerns, as well as a detailed account of the sort of mechanism they have in mind, see Mele and Robb 2003.} Of course, as is often the case in philosophical discourse, there is plenty of controversy to go around. Each of the various attempts to modify traditional Frankfurt scenarios faces its own set of objections challenging its ability to provide ensuring conditions that guarantee that an agent could not have done otherwise without playing a causal role in the agent’s actual behavior. I am not at present concerned with settling any of these disputes. My aim in this chapter, rather, is to show that even if we grant that traditional Frankfurt cases can be reformulated in a way that avoids the sting of the objection facing prior signs, these cases still fail to falsify PAP. It is to this task that I now turn.

\subsection*{2.4 For What is Jones Morally Responsible?}

Recall that a successful counterexample to PAP will be one in which (i) an agent S does something A, (ii) S is morally responsible for doing A, and (iii) S could not have done otherwise
than A.\textsuperscript{8} Furthermore, A must remain fixed in (i), (ii), and (iii). The thing that S does and for which S is morally responsible must be the very same thing that S could not have avoided doing. This much seems obvious.

But is it the fact that Jones sets fire to the post office for which he is morally responsible, or is it something else?\textsuperscript{9} Like all Frankfurt-type scenarios, Modified Post Office is set up in such a way that, regardless of whatever else Jones might do, the outcome is the same: The facts of the case guarantee that Jones will set fire to the post office. It is for this reason that Frankfurt claims Jones could not have done otherwise.

In this way, Frankfurt-type counterexamples (both in their modified and traditional forms) are structurally similar to the following case; call it “Box.”\textsuperscript{10} Imagine that an evil demon creates a world in which all that exists is a large box (roughly the size of a small studio apartment) containing a single human being, Jack. Further suppose that Jack and the box are each created in a way that makes it physically impossible for Jack ever to get out of the box. Unsurprisingly, Jack lives his entire life, and eventually dies, without ever setting foot outside the box.

Arguably, what has been said thus far is sufficient for precluding Jack from being morally responsible for the fact that he stays inside the box. After all, prior to his performing a single action, the stipulated facts of the case, over which he has absolutely no control, guarantee that Jack will forever stay in the box. It is, therefore, difficult to see how one would even begin to consider him morally responsible for this fact. To be clear, this is not to say that Jack could not be morally responsible for anything he does vis-à-vis his remaining in the box. Surely there are things for which Jack could, under the right circumstances, be morally responsible. For

\textsuperscript{8} Here, ‘A’ is meant merely as a stand-in for whatever it is for which one is morally responsible. It is left open whether ‘A’ should be understood narrowly to refer only to actions or more broadly to include such things as consequences of action, modes of action, and omissions, as well, perhaps, as beliefs and desires.

\textsuperscript{9} For ease of expression, I will sometimes refer simply to Jones’s setting the post office on fire. All such references, however, should be understood to apply both to his setting fire to the post office and to his decision to do so.

\textsuperscript{10} Readers will recognize Box as a straightforwardly souped-up version of John Locke’s (1690, bk. II, ch. xxi) locked room.
instance, not knowing that it is physically impossible, Jack might never have made any effort to get out of the box, despite believing that he had overriding moral reasons to do so. Here, however, what Jack would be morally responsible for is not the fact that he never makes it out of the box but, rather, the fact that he never even tries to get out—that is, not for his staying in the box simpliciter but, rather, for the particular way in which he stays in the box. And the same is true of modified Frankfurt cases, or so one might argue.\(^\text{11}\)

If this is correct, then, returning to Modified Post Office, what Jones is morally responsible for is setting the post office ablaze on his own, not for setting the fire simpliciter. The reason Jones is not morally responsible for torching the post office simpliciter is that things beyond his control have been set up in such a way as to ensure that, regardless of whatever else he does, whatever efforts he makes, Jones will set fire to the post office.\(^\text{12}\) The setup of this particular Frankfurt case can be seen as establishing the following disjunction: Either Jones will set fire to the post office on his own or Jones will set fire to the post office as a result of Black’s intervention. That the disjunction is true is stipulated up front, and there is nothing Jones can do to bring it about that the disjunction is false. Therefore, since the truth value of the disjunction is entirely out of his control, it would seem unreasonable to hold Jones morally responsible for it.\(^\text{13}\)

\(^\text{11}\) Perhaps some will worry that, unlike counterfactual interveners in Frankfurt scenarios, the box plays an actual role in Jack’s remaining inside it. It is important to note, however, that since Jack never even attempts to get out of the box, the nature of the box itself—in particular, those properties that make it impossible for Jack to break through—play no causal role in bringing it about that Jack acts as he does. Just as in Frankfurt-type counterexamples, Jack acts just as he would have had the box not been impossible to penetrate.

\(^\text{12}\) To be clear, the argument being made here is that, just as in Box, it is the conjunction of the fact that Jones’s circumstances ensure he will set fire to the post office together with the fact that his being in this situation was entirely beyond his control which prevents him from being morally responsible for setting the post office on fire simpliciter. Thus, the reasons offered here for denying Jones’s moral responsibility for setting fire to the post office simpliciter are ones that could plausibly be found to have force even for one not already committed to the truth of PAP. Even one who is not inclined to deny that Jones could be morally responsible for setting the post office on fire simpliciter merely on the grounds that circumstances ensured he would do so might nevertheless think this judgment warranted when it is added that his being in such circumstances was something over which he had absolutely no control.

\(^\text{13}\) Peter van Inwagen (1978) makes a similar point while arguing for a version of the flicker defense that focuses on the ability of agents in Frankfurt-type counterexamples to prevent certain particular events from occurring. Roughly, his argument is that Frankfurt scenarios guarantee the
What Jones is morally responsible for is determining which of the disjuncts it is that ends up making the disjunction true. If he enkindles the post office on his own, without any interference from Black’s device, he will bring it about that the first disjunct is true and the second one false. If, on the other hand, Jones does not set fire to the post office on his own, he will thereby make the first disjunct false, thus forcing Black’s device to come into play in order to make the disjunction true. Whether Black’s device comes into play is entirely contingent upon whether Jones makes the first disjunct true by torching the post office on his own. As in all Frankfurt cases, Jones makes the first disjunct true without any interference from Black or his device. And since Jones acts on his own, we are quick to say that he is morally responsible for what he does. Perhaps what Jones does, however, is not to make it the case that he sets fire to the post office, as Frankfurt supposes (just as Jack does not make it the case that he spends his entire life inside the box); for, again, that Jones will set fire to the post office is guaranteed up front, prior to anything Jones does. Instead, like Jack, what Jones does is bring it about that he sets fire to the post office *in the way that he does*—namely, on his own and without any interference from Black or his device. In this case, it is his setting the post office ablaze *on his own* for which Jones is morally responsible.  

As Frankfurt himself is quick to point out, however, it is well within Jones’s power to bring it about that he does not set fire to the post office on his own. He writes: “Of course it is in a way up to [Jones] whether he acts on his own or as a result of Black’s intervention. That depends upon what action he himself is inclined to perform. But whether he finally acts on his own or as a result of Black’s intervention, he performs the same action” (1969, p. 836). Thus, although Jones is morally responsible for setting fire to the post office on his own, it was obtaining of a state of affairs in which an event of a certain type occurs. Which particular event (of that type) will occur, however, remains entirely up to the agent. Assuming that agents are morally responsible for the occurrence of event particulars, rather than for the fact that a certain state of affairs was obtained, he argues, Frankfurt cases fail to falsify the view that moral responsibility requires alternative possibilities. John Martin Fischer (1986, pp. 180-1) contends that it is a weakness of van Inwagen’s strategy that its success depends on a contentious (and, in his mind, dubious) essentialist principle of event individuation, according to which $x$ is the same particular event as $y$ if and only if $x$ and $y$ have the same causal history. I take it to be an advantage of the versions of the flicker strategy developed in this chapter that they remain neutral on the question of how actions and events are to be individuated.

---

14 Marjory Naylor (1984) has also argued for this point.
possible for him not do so. And while it was not within his power to prevent himself from setting
fire to the post office, it is not this fact for which he is morally responsible. Insofar as this is
correct, like their predecessors, modified Frankfurt-type scenarios fail to provide us with cases of
agents who are morally responsible for doing something they could not have avoided doing, and
so fail to provide us with good grounds for rejecting PAP.

2.5 An Artificial Separation?
The kind of strategy for resisting Frankfurt-type counterexamples offered above has prompted
some concerns regarding the connection between moral responsibility for A-ing on one’s own
and moral responsibility for A-ing simpliciter. Robert Kane has argued that, although it is true in
Frankfurt scenarios that Jones could not have done otherwise than A, “it is implausible to add
that Jones is therefore not responsible for doing A even when he does A on his own” (1996, p.
41). The reason for this, he says, is that “This step too artificially separates responsibility for
doing-A-on-one’s-own from responsibility for doing A. In general, if we are responsible for
doing something on our own, we are responsible for doing it. And this seems also to hold in
Jones’s case” (p. 41).

It may very well be true that, in general—that is, in normal cases—responsibility for A-
ing on one’s own and responsibility for A-ing are all of a piece. But Frankfurt cases plainly are
not normal cases. Normally, no one is following us around, monitoring our blushing activity, or
implanting special devices in our brains in order to force us to act in a particular way in the event
that we will not do so on our own. Even if Kane is correct in claiming that, in general, we tend to
think that if a person is morally responsible for A-ing on her own she is also morally responsible
for A-ing, it is entirely plausible that this is (at least partly) because in normal cases we also
think that had a person not A-ed on her own she would not have A-ed at all. Such, of course, is
not the case in Frankfurt scenarios. Thus, we should not assume that what holds for cases of
moral responsibility in general will also hold for extremely unusual cases of this sort.15

15 Similarly, we might also think that, in general, a person is morally responsible for A-ing only
if he is also morally responsible for A-ing on his own. Here, too, however, we can imagine cases
where this generalization fails to hold true. Suppose, for example, that, wanting to burn down the
post office but fearing he lacks the nerve to go through with it on his own, Jones hires Black to
use his device to compel him to set the post office on fire. In this case, it seems pretty clear that
Depending on the peculiarities of a given situation, it may be that moral responsibility for A-ing on one’s own and moral responsibility for A-ing sometimes come apart. Consider the following case; call it “Witch Trial.” Nelda has been falsely accused of practicing witchcraft in seventeenth century New England, and a judge has ordered that she be burned at the stake. In addition, thinking it particularly fitting for witches to die by their own hands, the judge has ordered that Nelda be hypnotized into setting the blaze herself. On Monday morning, the day of her execution, one of Nelda’s friends manages to smuggle a small bottle of fast-acting, deadly poison into the cell where she is being kept, so that she may avoid the excruciating pain of burning alive. If Nelda takes the poison, she will die quickly and painlessly. If, on the other hand, she does not take the poison, then she will be hypnotized into setting herself on fire and will burn to death at the stake, in accordance with the judge’s orders. As it happens, Nelda takes the poison and dies a quick, painless death alone in her cell.

What should we say about Nelda? It certainly seems plausible to suppose that Nelda is morally responsible for killing herself on her own, which is something she could have avoided. It is also clear that it was not possible for Nelda to avoid killing herself altogether. The crucial question here, however, is whether Nelda is also morally responsible for the fact that she kills herself simpliciter. As Kane would have it, Nelda is indeed responsible for the fact that she kills herself simpliciter since she kills herself on her own. If an agent “does A-on-his-own,” he says, “there is no good reason to say that [he] is not responsible for doing A” (p.41). But given the special circumstances of Nelda’s situation, this conclusion may strike some as somewhat counterintuitive.

According to at least one way of thinking about these things, Nelda was not the one who brought it about that she ended up killing herself (or even that she killed herself on Monday) and so is not morally responsible for this fact. Rather, it is the judge, not Nelda, who brought it about that she would kill herself on Monday, which he did long before Nelda even began to consider whether to drink the poison (and so kill herself on her own). And while it was indeed up to Nelda Jones is (at least partly) morally responsible for torching the post office, even though this is not something he does (or is morally responsible for doing) on his own.

The claim that Nelda is morally responsible for drinking the poison and so killing herself on her own should not be taken to imply that she is blameworthy (or praiseworthy) for what she does. No assumption is being made about the moral value of her action.
to decide whether she would die quickly and painlessly as a result of drinking the poison or slowly and painfully at the stake, it was not up to her to determine whether she would kill herself (on Monday). That had already been guaranteed. Indeed, that Nelda killed herself (on Monday) was simply the unavoidable consequence of the situation in which she found herself, the creation of which was entirely beyond her control. Thus, there is something a bit off-putting about the suggestion that Nelda is morally responsible for this fact.

To be sure, there is a sense in which it seems quite natural in cases of this sort (where Nelda kills herself on her own) to say that Nelda is responsible for her suicide. But whether this serves to bolster the case against PAP will depend on what ‘her suicide’ is supposed to pick out. If it is meant to refer to the actual event of Nelda’s killing herself *as it happened* (namely, on her own by drinking the poison), then this certainly appears to be something for which she is morally responsible. However, since it is also true that she could have avoided killing herself in this way, Nelda’s responsibility for this poses no particular threat to PAP. If, on the other hand, ‘her suicide’ is intended to refer simply to the fact that she killed herself (or that she killed herself on Monday)—a fact which was guaranteed by factors wholly beyond her control and which would have been true even if she had refrained from killing herself on her own—then there appear to be reasonable grounds for resisting the claim that this is something for which Nelda is morally responsible. Thus, though there may be a sense in which it is natural to say that Nelda is morally responsible for her suicide, this does not necessarily amount to the claim that she is morally responsible for the fact that she killed herself simpliciter.

Even supposing Kane is correct, then, in maintaining that, generally, if one is morally responsible for A-ing on one’s own one is also morally responsible for A-ing, this may not always hold true. What cases like Witch Trial seem to suggest is that sometimes agents are not morally responsible for the fact that they A-ed even when they are morally responsible for the fact that they A-ed on their own. When circumstances beyond an agent’s control guarantee that she will act in a certain type of way (whether on her own or not), just as they do in Frankfurt-type counterexamples, it seems unreasonable to maintain that the agent can nevertheless be morally responsible for the fact that an action of that type is performed by her.17

---

17 A referee for *Philosophical Studies* points out that, assuming Nelda is aware that her sentence includes lighting her own pyre and that she will therefore kill herself either way, her situation is epistemically different from that of agents in Frankfurt-type counterexamples. As a result, the
Still, Kane is surely correct in noting that there is something strange about the view that one can be morally responsible for A-ing on one’s own and yet not morally responsible for A-ing, if for no other reason than that one’s having A-ed on one’s own appears to be sufficient for one’s having A-ed. Indeed, it is difficult to see how this would not be the case. Thus, even if the judge’s orders were sufficient to bring it about that Nelda kills herself, it may to some seem more natural to view this as an instance of causal overdetermination rather than deny that by killing herself on her own Nelda also brought it about that she killed herself simpliciter. Others may wish to deny that the judge’s orders were in any way causally efficacious. Though the judge’s orders guaranteed that Nelda would kill herself, it might be thought, they did not do so causally. It was Nelda’s drinking the poison, and not the judge’s orders, that brought it about that she killed herself on Monday. Either way, it is safe to assume that the claim that Nelda is not at all morally responsible for the fact that she killed herself, though she is morally responsible for the fact that she killed herself on her own, is, even if defensible, likely to be met with some resistance. It would be preferable, then, if defending PAP against Frankfurt-type counterexamples did not require commitment to this claim. And it seems that it does not. In the following section I suggest that a much more modest claim will turn the trick.

actual sequence leading up to Nelda’s taking the poison includes deliberation only over whether to kill herself on her own, not deliberation over whether to kill herself simpliciter. “This,” the referee suggests, “might arguably isolate her moral responsibility for killing herself on her own from her moral responsibility for killing herself simpliciter, in a way that wouldn’t happen in the typical Frankfurt-type counterexample, where the agent is unaware of the counterfactual intervener.” It is unclear, however, exactly why this difference in deliberative content (rather than the fact that her being in this situation was entirely beyond her control) would isolate Nelda’s responsibility for killing herself on her own from responsibility for killing herself simpliciter. Suppose someone cuts in front of me while I am waiting in line at the grocery store, and I immediately and non-actionally come to acquire an intention to hit him in the back of the head. I do not deliberate at all about whether to hit him in the back of the head since the role of deliberation is to resolve practical uncertainty about what to do, and I am not the least bit uncertain. Indeed, my only uncertainty in this case concerns which of the items in my basket to hit him with: the wine bottle or the frozen chicken. After brief deliberation, I quickly settle on the wine bottle and proceed to exact vengeance. It certainly seems that I am morally responsible both for hitting the person in the back of the head and for hitting him with a wine bottle, even though the actual sequence of events does not include deliberation over whether to hit him, only deliberation over what item to use to do so. Thus, the mere fact that an agent does not deliberate over whether to do a thing before doing it does not seem sufficient to preclude his being morally responsible for having done it.
2.6 A More Modest Approach

Perhaps the lesson from cases like Box and Witch Trial is not that agents in Frankfurt-type counterexamples (as well as other similar scenarios) are not at all morally responsible for A-ing, even though they are responsible for A-ing on their own, but merely that their responsibility for A-ing derives from their responsibility for A-ing on their own. On the face of it, this certainly seems to have the ring of plausibility. After all, it is only because Nelda kills (and is morally responsible for killing) herself on her own that we are even tempted to say that she is responsible for killing herself. If a person is not morally responsible for doing a thing on her own, it is difficult to see how she could be morally responsible for doing it simpliciter (assuming she did not on her own bring it about that she would do it but not on her own; see n. 12).

In addition to being plausible, the view that one’s responsibility for A-ing derives from one’s responsibility for A-ing on one’s own looks to be something even opponents of PAP can (and do) readily endorse. The primary reason offered by proponents of Frankfurt-type counterexamples for thinking that agents in these cases are morally responsible for what they do is the fact that they do it on their own, just as they would have had no counterfactually intervening mechanism been present. Indeed, when discussing his original case, Frankfurt himself argues that although Jones “has no alternative but to do what Black wants him to do…If he does it on his own…his moral responsibility for doing it is not affected by the fact that Black was lurking in the background with sinister intent, since this intent never comes into play” (1969, p. 836). Now, while not strictly entailed by what Frankfurt says, the clear implication here is that had Jones not acted on his own he would not be morally responsible for what he did. Thus, the claim (highlighted by cases like Box and Witch Trial) that agents’ responsibility for what they do in Frankfurt-type cases derives from their responsibility for doing it on their own does not look as though it should be particularly controversial.

Other considerations seem also to support the view that agents’ responsibility for what they do in Frankfurt-type scenarios (and perhaps normally) derives from their responsibility for doing it on their own. It is commonly held that responsibility for overt physical actions (such as raising one’s arm) derives from responsibility for certain mental actions (such as deciding to raise one’s arm), much in the same way that responsibility for the consequences of one’s actions derives from one’s responsibility for the actions themselves. On this view, if one is morally responsible both for deciding to raise one’s arm and for raising one’s arm, it is the former for
which one is basically (i.e., directly or non-derivatively) morally responsible, whereas one is only derivatively (or indirectly) responsible for the latter. This variation in the kind of responsibility agents bear for certain mental actions as compared to overt physical actions is explained by a difference in control. The successful performance of even simple overt actions like raising one’s arm is much more vulnerable to environmental impediments than that of mental actions like deciding to raise one’s arm. It is because of the increased control agents have over certain mental actions, such as decisions, that it is these, rather than overt physical actions, for which they are basically morally responsible.

In Frankfurt scenarios, however, the situation is a little different. In Modified Post Office, for instance, Jones has the same amount of control over whether he will set fire to the post office as he does over whether he will decide to set fire to the post office—namely, none. That Jones will decide to set fire to the post office and that he will do so have been guaranteed, and there is nothing he can do to prevent either of these things from occurring. But even in Frankfurt cases, whether Jones will decide on his own to set fire to the post office and whether he will set fire to the post office on his own remain entirely up to him. Thus, the same reasons typically offered in support of the view that one’s responsibility for performing some overt action A derives from one’s responsibility for some mental action, such as deciding to A, equally suggest that, at least in Frankfurt scenarios, one’s responsibility for A-ing derives from one’s responsibility for so deciding on one’s own (for which one is basically morally responsible).

Now, then, how is this supposed to aid defenders of PAP in answering the challenge posed by Frankfurt-type counterexamples? To see this, we should begin by noting that, as traditionally formulated, PAP is notoriously ambiguous. Frankfurt states it this way:

18 Insofar as this correct, some might want to maintain that it is deciding on one’s own to do a thing, rather than so deciding simpliciter, for which one is always basically morally responsible, not just in Frankfurt-type scenarios (and other similar cases). No doubt the congruity of such a view is appealing. Depending on how closely one takes the distinction between the objects of basic and derivative moral responsibility to track differences in control, however, one might hold that considerations of this sort ought to outweigh considerations of symmetry. By these lights, if in typical (non-Frankfurtian) cases the control agents have over whether they will decide to do a thing simpliciter matches the control they have over whether they will so decide on their own, then, unlike in Frankfurt cases, there might not be any good reason to deny that they can be basically morally responsible both for deciding on their own and for deciding simpliciter. I take no stand here on this issue.
PAP: A person is morally responsible for what he has done only if he could have done otherwise. (1969, p. 829)

How should we interpret this? At the heart of this principle, plainly enough, is the claim that some sort of possibility of doing otherwise is a necessary condition for moral responsibility. Beyond this, things are less clear. Is PAP a thesis about moral responsibility for actions only, or does it extend to the consequences of one’s actions as well? Does it apply to things for which one is only derivatively morally responsible, or is it restricted to cases of basic (or non-derivative) responsibility? What is it exactly that one must have been able to do other than, and what does doing otherwise involve? Is it sufficient for one’s having been able to do otherwise that it was possible for one not to have done what one did, or does this require something much stronger, such as the ability to have performed an alternative intentional action? There is much it does not say. Depending on how one answers these and other related questions, one will end up with significantly different versions of the same general principle, some of them more plausible and less vulnerable to falsification by Frankfurt cases than others. By removing some of these ambiguities we will be in a better position to evaluate the extent to which Frankfurt-type counterexamples are successful in falsifying the basic intuition underlying PAP, rather than merely exploiting a semantic weakness in the way this principle has traditionally been formulated.

Let us suppose, as is generally accepted, that PAP is a thesis about basic moral responsibility, which we can specify in the following way:

PAPb: A person is basically morally responsible for what he has done only if he could have done otherwise.

Do Frankfurt-type counterexamples like Modified Post Office succeed in falsifying PAP when understood in this way? Whether PAPb is falsified by cases of this sort will depend on how we answer some of the other questions just mentioned. First, regarding what, exactly, must Jones have been able to do otherwise? That is, precisely to what does the requirement that Jones could have done otherwise refer? According to PAPb, what Jones must have been able to do other than
is that (whatever it is) for which he is basically morally responsible. The referent of the former is fixed by that of the latter. Thus, specifying the object of basic moral responsibility in Modified Post Office will also serve to identify that regarding which Jones must have been able to do otherwise, according to PAPb.

It is at this point that the more modest approach begins to yield fruit for the defender of alternative possibilities. If the foregoing is correct, what Jones is basically morally responsible for is deciding on his own to set fire to the post office. Again, this is not to deny that Jones is morally responsible for deciding (simpliciter) to set the post office on fire but merely to maintain that whatever responsibility he bears for so deciding simpliciter derives from his responsibility for doing so on his own. According to PAPb, then, Jones will be basically morally responsible for deciding on his own to set fire to the post office only if he could have done otherwise than decide on his own to set fire to the post office.\textsuperscript{19} Thus, Modified Post Office will be successful in falsifying PAPb only to the extent that it is true that Jones is basically morally responsible for deciding on his own to set fire to the post office and yet false that Jones could have done otherwise than decide on his own to set the post office on fire. As we have seen, however, though Jones could not have done otherwise than decide to set fire to the post office, whether he does so on his own remains entirely up to him. Modified Post Office fails, then, to provide us with a genuine counterexample to PAPb—one in which a person could not have done otherwise than for which he is basically morally responsible.

\section*{2.7 The Robustness Objection}

The primary objection to the sort of flicker strategy I have been defending comes from John Martin Fischer. He argues that even if it is possible for agents in Frankfurt-type counterexamples not to decide on their own to act as they do, this is no help to defenders of PAP (even when

\textsuperscript{19}Though I maintain that what agents in Frankfurt-type counterexamples are basically morally responsible for is deciding on their own to act as they do, PAPb itself remains neutral on the issue of whether the class of objects of basic moral responsibility ought to be restricted to certain modes of action (such as doing a thing on one’s own), certain types of mental action (such as deciding), or whether it should include such things as overt physical actions, consequences of actions, beliefs, and desires. In this way, and perhaps unlike some other principles offered by flicker theorists, PAPb is able to, as Seth Shabo puts it, retain “the sheen of an all-purpose principle, one with a life in and outside of Frankfurt-style scenarios” (2007, p. 69). Cf. Timpe 2006.
understood as PAPb) because “this alternative possibility is not sufficiently robust to ground the relevant attributions of moral responsibility” (1994, p. 140). In order for the existence of alternative possibilities in Frankfurt-type counterexamples to neutralize the threat to PAP, he says, the alternatives must be able to account for why an agent would, in virtue of possessing these alternatives, be thereby morally responsible for what he does. He writes:

> even if the possible event at the terminus of the alternative sequence (in the case of Jones and Black) is indeed an alternative possibility, it is highly implausible to suppose that it is in virtue of the existence of such an alternative possibility that Jones is morally responsible for what he does. I suggest that it is not enough for the flicker theorist to analyze the relevant range of cases in such a way as to identify an alternative possibility. Although this is surely a first step, it is not enough to establish the flicker of freedom view, because what needs also to be shown is that these alternative possibilities play a certain role in the appropriate understanding of the cases. That is, it needs to be shown that these alternative possibilities ground our attributions of moral responsibility. (p. 140)

According to Fischer, then, defending PAP (or some version of it) against the onslaught of Frankfurt-type counterexamples requires demonstrating not only that agents in these cases could have done otherwise than that for which they for which they are (basically) morally responsible but that the possibilities that remain open to them are at least partly constitutive of their moral responsibility for what they do. As he later puts it, “the flicker theorist must claim that the addition of the sort of possibility he has identified would transform a case of lack of responsibility into one of responsibility” (p. 141).

But why is it that the possibility of refraining from deciding on one’s own to do what one does in Frankfurt-type counterexamples is not sufficiently robust to ground the relevant attributions of moral responsibility? To see this, Fischer asks us to consider traditional Frankfurt-type cases like the original version of Post Office, where Black’s intervention is contingent on whether Jones displays some sort of sign (such as a blush or a twitch) indicating that, left on his own, he will not freely do what Black wants him to do (pp. 144-5). Fischer is willing to grant that it remains open to agents in these cases not to decide on their own to act as they do but, rather, to so decide as a result of Black’s intervention, which they can bring about by displaying
(or failing to display) the relevant sign. By hypothesis, however, the nature of these signs is such that they are involuntary, and so lie beyond the scope of the agent’s direct voluntary control. Thus, even though it remains open to agents in these cases to act not on their own but as a result of Black’s intervention, whether they do so is not something over which they have any significant kind of control. It is because the addition of alternatives of this sort in no way adds to the control that the agent has over what he does, Fischer maintains, that this cannot plausibly be held to contribute to the agent’s moral responsibility for what he does, and so cannot be used to defend PAP against Frankfurt-type attack.

The first thing to note about Fischer’s objection to the flickers-of-freedom defense of PAP is just how heavily it relies on the presence of some involuntary sign that indicates what an agent will freely do if left to act on his own. It is this very feature of traditional Frankfurt-type counterexamples, however, which has come under heavy attack in recent years, prompting the creation of modified Frankfurt cases that do not rely on signs of this sort. The significance for the flicker strategy of the rejection of prior signs in the construction of modified Frankfurt-type counterexamples is that it ensures that up until the moment of decision it is possible for agents in these cases to realize one of the alternative possibilities that remain. Because these possibilities are available up until the moment of choice (the locus of free action) and within the range of agents’ voluntary control, it is far from obvious that they are insufficiently robust to help ground ascriptions of moral responsibility—at least, not for the reasons Fischer provides. Thus, regardless of how successful Fischer’s objection is in attenuating the flicker defense of alternative possibilities against traditional Frankfurt-type counterexamples, when employed in defense of modified Frankfurt scenarios, it falls flat.

Even in the context of traditional Frankfurt-type scenarios, however, Fischer’s attempt to throw water on the flicker strategy is puzzling. Among other things, it is unclear why we should think that defending PAP against Frankfurt-style attack requires demonstrating that the alternatives that remain open to agents in these cases are robust in the way Fischer suggests. PAP maintains that agents are (basically) morally responsible for what they have done only if they could have done otherwise. Proponents of Frankfurt-type counterexamples (both in their modified and traditional forms) argue that these cases falsify PAP. The aim of the flicker strategy is to rebuff this attack, which it does by arguing that, since agents in these cases could have done otherwise than that for which they are (basically) morally responsible, these cases are not
genuine counterexamples to PAP. But, then, the demand at this point that flicker strategists show not only that agents in these cases do in fact possess alternative possibilities—that is, not only that Frankfurt-type scenarios fail to falsify PAP—but that these alternatives are relevant to agents’ moral responsibility for what they do in these cases seems misplaced. In effect, this is just to demand that flicker defenders show that PAP is true. Although it is in defense of PAP that the flicker strategy is employed, it is not a defense in the sense of a full-fledged *apologia*; rather it is a defense of the principle *against* Frankfurt-type attack.

Perhaps the biggest problem with the robustness objection, then, is that it seems to run together two separate questions. First, do Frankfurt-type counterexamples falsify the view that moral responsibility for what one does requires alternative possibilities? Second, what does the mere possession of alternative possibilities add to an agent who already satisfies all the other necessary conditions for moral responsibility save the ability to do otherwise? These questions, though related, are in fact quite distinct, and answering the former does not require answering the latter. Whereas the second question has to do with whether alternative possibilities are in fact necessary for moral responsibility, the first is concerned merely with whether Frankfurt-type counterexamples show that they are not. One may wonder, of course, what the mere possession of alternative possibilities adds to an agent who already satisfies all the other necessary conditions for morally responsible agency save the ability to do otherwise. This is a fair question. It is important to see, however, that it does not directly bear on the issue of whether Frankfurt-type counterexamples are successful in falsifying PAP (particularly when understood as a thesis about basic moral responsibility).

### 2.8 Conclusion

In this chapter I have tried to show that even when Frankfurt-type counterexamples have been reformulated in a way that no longer begs the question against incompatibilists by relying on prior signs that presuppose a deterministic relation between an agent’s free action and its antecedents, they are still unable to do the work set out for them. My direct focus has been on considering the particular kind of modified Frankfurt-type scenario represented in Modified Post Office, offered by Mele and Robb. Setting aside worries about occurent causal preemption, and so granting that Jones could not have done otherwise than perform the action he performs, given that PAP is properly understood as a thesis about basic moral responsibility (as characterized in
PAPb), Modified Post Office will be successful in showing PAP to be false only if Jones is basically morally responsible for deciding to set fire to the post office simpliciter. Considerations arising from cases like Box and Witch Trial, however, appear to provide plausible grounds for denying that Jones is at all morally responsible for deciding to set fire to the post office simpliciter, let alone basically morally responsible. Moreover, even assuming this is incorrect, and that Jones is morally responsible both for deciding on his own to set fire to the post office and for deciding to set fire to the post office simpliciter, it is the former, not the latter, for which he is basically morally responsible. Either way, PAP can accommodate Modified Post Office.

Of course, Mele and Robb’s is not the only kind of modified Frankfurt scenario on offer (see n. 4). The controversy surrounding determinative prior signs has given rise to a number of modified Frankfurt cases, which differ (at times significantly) on the range of alternative possibilities they eliminate as well as on the nature of the mechanisms they employ to do so. Thus, objections that are effective against some cases will not necessarily work when mounted against other cases that are structurally dissimilar.

As I stated earlier, however, nothing about the particular versions of the flicker defense developed in this chapter hinge on anything unique to Modified Post Office. Instead, they capitalize on two features of this case that are essential to any Frankfurt-type scenario. First, the agent acts on his own. That agents in these cases act on their own, just as they would have had no ensuring mechanism been in place, is what is supposed to drive the intuition that they are morally responsible for what they do, without which there is no threat to PAP. Second, the ensuring mechanism is not just for show. Rather than being superfluous, the ensuring mechanisms (whatever forms they take) lie at the very heart of Frankfurt cases, as they are supposed to be what makes it true that the agents could not have done otherwise than perform the actions they perform, which they do by standing ready to force the agents to behave in the desired ways should the agents not do so on their own. Most importantly, not doing so on their own is a genuine alternative possibility in all Frankfurt-type scenarios (and in modified cases is an alternative that lies safely within the scope of agents’ direct voluntary control). Indeed, it is precisely because it is possible for agents in these cases not to perform the relevant actions on their own that the ensuring mechanisms are there in the first place (otherwise they would be extraneous). And it is because this alternative possibility is ineliminable that the particular
versions of the flicker defense advanced here can be brought to bear against any modified Frankfurt-type case, not just Modified Post Office.\textsuperscript{20}

Insofar, then, as either version of the flicker strategy defended in this chapter is successful, it shows not only that PAP has yet to be falsified by any of the modified Frankfurt-type cases currently on offer but that cases of this sort are in principle incapable of falsifying PAP. Frankfurt-defenders wishing to overcome the kinds of flicker defense offered here will need to do one of two things: Either show that agents in these cases are basically morally responsible not (only) for deciding on their own to act as they do but for so deciding simpliciter, or else show both that the (alternative) possibility of refraining from deciding on their own to act as they do is not sufficiently robust (despite the fact that this lies within their direct voluntary control in these cases) and that defending PAP against falsification by Frankfurt-style attack in fact requires that this alternative be robust in the sense described. Unless this can be done, proponents of the view that moral responsibility requires alternative possibilities have nothing to fear from modified Frankfurt-type counterexamples.

\textsuperscript{20} Thanks to an anonymous referee for Philosophical Studies for encouraging me to clarify this point.
CHAPTER THREE

THE LIMITS OF LIMITED-BLOCKAGE
FRANKFURT-STYLE CASES*

3.1 Introduction

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. Perhaps the greatest challenge currently facing this seemingly intuitive principle comes from a class of putative counterexamples developed by Harry Frankfurt (1969), commonly known as Frankfurt-style cases. These are meant to be cases in which agents are clearly morally responsible for having done things they could not have avoided doing. Philosophers employing Frankfurt-style cases to challenge PAP have mostly sought to construct scenarios that eliminate as many of an agent’s alternatives as possible—and all alternatives at the moment of action, within the agent’s control—but without causally determining the agent’s actions (lest this call into question the agent’s moral responsibility). One of the chief difficulties for this traditional approach is that the closer one gets to eliminating absolutely all alternative possibilities the more it appears that agents’ actions in these cases are causally determined. “Limited-blockage” versions of these cases are intended to sidestep this worry by blocking all and only those alternatives that are intrinsically relevant to moral responsibility (“robust alternatives”) while leaving open all other alternatives, including a significant range of alternatives that are within the agent’s voluntary control at the moment of action.21 I argue that, owing to the fact that omissions (and not just actions) are capable of constituting robust alternative possibilities, limited-blockage cases cannot avoid collapsing into

* Forthcoming in *Philosophical Studies*. Printed by permission of the publisher, for which I am grateful.

21 The term ‘robust alternatives’ was coined by John Martin Fischer (1994, pp. 137-47) to refer to alternative possibilities that are capable of grounding ascriptions of moral responsibility. In order to be robust, an alternative must figure into an explanation of why an agent is morally responsible for what she actually does.
the more traditional sort of Frankfurt-style case to which they are meant to be an alternative and so are vulnerable to the very same concerns they are meant to avoid.

3.2 The Trouble with Traditional Frankfurt-Style Cases
A genuine counterexample to PAP will be a case in which an agent is morally responsible for performing some action she could not have done otherwise than perform. Because PAP is generally understood to be a thesis about basic (i.e., direct, or non-derivative) moral responsibility, the relevant actions in these cases tend to be mental acts of deciding, which are commonly taken to be the things for which we are basically morally responsible and from which our responsibility for overt physical actions (and their consequences) derives. To ensure that agents cannot do otherwise than decide to act as they actually do, typical Frankfurt-style cases include the presence of a counterfactual intervener, someone (or something) who stands ready and able to force the agent to decide to act in the desired way in the event that the agent is not going to do so on her own. Crucially, however, the counterfactual intervener’s presence is unknown to the agent and plays no role in the agent’s actual behavior, lest it undermine the intuition that the agent is morally responsible for what she does.

The tricky part for cases of this sort comes in identifying the precise conditions under which the counterfactual intervener would intervene. If the counterfactual intervener has to wait to see how the agent actually decides, then the agent retains the ability to do otherwise and PAP is unthreatened. In order to guarantee that the agent will decide to act in the desired way without actually compelling the agent’s behavior, the counterfactual intervener must have some way of knowing in advance what the agent will do if left to act on her own. Toward this end, traditional

---

22 Counterexamples to PAP need not be restricted to cases involving actions. Since PAP is generally taken to apply to both actions and omissions, a case in which an agent is genuinely morally responsible for omitting to perform some action she could not have done otherwise than omit to perform would also serve as a counterexample to this principle. Here, following custom, I focus on cases involving moral responsibility for actions (as opposed to omissions) merely for the sake of simplicity.

23 An agent is basically morally responsible for some action she performs when her responsibility for that action is not inherited from (or in virtue of) her responsibility for other actions. Moral responsibility that is inherited in this way is derivative, or indirect (see Mele 2006, p. 86). For arguments that agents can be basically morally responsible for things other than certain mental actions, see Adams 1985 and Smith 2005, 2006.
Frankfurt-style cases posit the existence of a *prior sign*—something before the moment of decision and beyond the agent’s voluntary control (such as a blush or a cluster of neural firings) that indicates how the agent would freely act were no intervention to occur.

The trouble with including indicators of this sort is that they are capable of doing the work set out for them in traditional Frankfurt-style cases only by presupposing a deterministic relation between an agent’s free action and its causal antecedents—a relation incompatibilists flatly reject (see Widerker 1995; Kane 1996; Ginet 1996). This problem can be seen as posing the following dilemma: Either the occurrences of these prior signs (or that which they indicate) make it merely probable that agents in these cases will decide to act in the desired ways, or they deterministically lead to the agents so deciding. If the occurrences of these signs make it merely probable that the agents will decide to act in the desired ways, then it remains causally open to them not to decide to act in this way after all. Insofar as this is the case, however, agents will continue to possess alternative possibilities with respect to their actions, and so their bearing moral responsibility for these actions is perfectly compatible with PAP. If, on the other hand, the occurrences of these signs (or that which they indicate) will deterministically lead to the agents deciding to act in the desired ways, then it will no longer remain causally open to the agents not to decide to act as they do. But since having one’s actions causally determined precludes one from being morally responsible for them, according to incompatibilists, that agents in these cases are nevertheless morally responsible for what they do cannot be asserted without begging any questions. 24 Either way, then, traditional Frankfurt-style scenarios fail to constitute genuine counterexamples to PAP. 25

---

24 For arguments to the effect that there is in fact nothing illicit in the way traditional Frankfurt-type counterexamples make use of (determinative) prior signs, see Fischer 1999b; Haji and McKenna 2004.

25 It should be noted that, relying as it does on the decidedly incompatibilist intuition that determinism and moral responsibility are mutually exclusive, this is not a line of argument available to compatibilist defenders of PAP. Compatibilists who wish to resist the conclusion of Frankfurt’s argument cannot do so on the grounds that it is question-begging to assert that agents can be morally responsible for their actions despite their being causally determined.
3.3 The Limited-Blockage Strategy

This particular line of objection has prompted numerous attempts to modify traditional Frankfurt-style cases in ways that will enable them to do the work set out for them by Frankfurt without the use of prior signs. Michael McKenna has argued that the trouble with most (if not all) of these attempts is that, like traditional Frankfurt cases, they seek to eliminate all an agent’s alternatives at (but not prior to) the moment of choice, the locus of free action. This is both problematic and unnecessary. It is problematic, he says, because ruling out absolutely all alternatives at (or just before) the moment of action, within the scope of an agent’s voluntary control, “comes dangerously close to making that problematic deterministic assumption” (2003, p. 206). Furthermore, it is unnecessary because, contrary to what these attempts seem to presuppose, not all alternatives within an agent’s voluntary control are robust. A better tack, he says, would be to eliminate all and only those alternatives that are robust while leaving open all other alternatives, including a wide range of non-robust alternatives at the moment of action, within the agent’s voluntary control.

In this vein, McKenna offers the following limited-blockage Frankfurt-style case, which he calls 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

---

26 See, for example, Haji 1998; Hunt 2000; McKenna 2003; Mele and Robb 1998, 2003; Pereboom 2000, 2009; Stump 1996.
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
… [A]s 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. (pp. 209-210)

Since, we are to suppose, the lesion on Casper’s brain plays no causal role in his decision to
press the Bad button and he acts just as he would have had the lesion not been present, there
appears to be no reason to deny that he is morally responsible for what he does. Grant that this is
correct.

As McKenna is quick to point out, Brain Malfunction is not, nor is it intended to be, a
counterexample to PAP. Even assuming Casper is morally responsible for deciding to press the
Bad button, this is not something he had to do; a wide range of alternatives remained open to
him. These included continuing to deliberate about what to do; ceasing to deliberate and turning
his attention elsewhere; even deciding that, whatever else he did, he most definitely would not
press the Bad button. Indeed, the only thing Casper could not have done, McKenna says, is
decide to press the Good button (p. 210).

How, then, is this case supposed to threaten PAP? The answer, it seems, is that it is not.
The aim of Brain Malfunction (and limited-blockage cases in general, presumably) is to function
as a counterexample not to PAP but to another principle, which McKenna calls the Principle of Morally Significant Alternatives for blame (PSA):

PSA: An agent S is morally blameworthy for performing action A at t only if she had within her control at t performing an alternative action B such that (1) performing B at t was morally less bad than performing A at t, and (2) it would have been reasonable for S to have considered performing B at t as an alternative to performing A at t given S’s agent-relative deliberative circumstances. (p. 209)

McKenna argues that it is this principle, not PAP, which ought to serve as the incompatibilist demand for alternative possibilities (p. 210). The reason for this, he says, is that PAP is “too inclusive,” as it “can be satisfied even if the only options available to an agent are morally (that is, deliberatively) insignificant ones” (pp. 208-209). McKenna’s overall approach, then, is to begin by arguing that PAP ought to be rejected in favor of PSA and then proceed to use cases like Brain Malfunction to show that PSA is false.

Before considering whether Brain Malfunction is successful in demonstrating that moral responsibility does not require robust alternative possibilities (of the sort identified in PSA), a few points bear mentioning. First, dialectically speaking, there is something a bit curious about telling proponents of a certain principle (e.g., PAP) that they ought to reject it in favor of some other principle (e.g., PSA) while at the same time purporting to offer counterexamples that show the second principle to be false. Proponents of PAP will surely be forgiven if they are a little wary of this admonition and perhaps somewhat reluctant to follow this advice. At the very least, it seems, they should not do so in the absence of a pretty compelling argument—which brings us to the next point.

The only reason McKenna offers in support of the claim that proponents of PAP ought to reject this principle in favor of PSA is that PAP is “too inclusive,” which is to say that it is too easily satisfied. This is a puzzling criticism of PAP. Why exactly is it a problem that the alternative possibilities requirement stated in PAP is easy to satisfy? That is, why think it is too

---

27 Presumably McKenna would also say that compatibilists who maintain that moral responsibility requires the ability to do otherwise should abandon PAP in favor of PSA. Here, however, his focus is restricted to incompatibilist proponents of PAP.
easy to satisfy? Perhaps this criticism would make sense if PAP purported to offer conditions that are sufficient for moral responsibility. But PAP does not do this. It does not state that the ability to do otherwise is a sufficient condition for being a morally responsible agent, nor even that it is sufficient for having the freedom that moral responsibility requires. Rather, it merely offers a necessary condition for moral responsibility. What does it mean, then, to say that a necessary condition is too easy to satisfy? Stronger requirements always entail weaker ones. If it is true that moral responsibility requires \(x, y\), and \(z\), then it is also true that moral responsibility requires \(x\). Consequently, it is difficult to know what to make of the claim that one ought to accept PSA instead of PAP. PSA entails PAP.\(^{28}\) So, anyone who accepts the former is thereby logically committed to the latter.

Motivating McKenna’s claim here, it seems, is the intuition that meeting the conditions for morally responsible agency must involve more than merely possessing the ability to do otherwise. Assume this is correct. It is still unclear why this poses any sort of problem for PAP. Since PAP purports to offer only a necessary condition for moral responsibility, it is entirely consistent with the view that moral responsibility requires more than the ability to do otherwise. In fact, the claim that there are additional conditions that must be satisfied in order for an agent

\(^{28}\) Strictly speaking, given that PSA is restricted to cases of blameworthiness, it entails only a version of PAP that is similarly restricted in scope. According to McKenna, this is done only for ease of presentation. Presumably, then, when McKenna says that “it is PSA and not PAP that ought to serve as the incompatibilist demand for alternative possibilities,” he means to refer to a broadened version of PSA, one with application to all cases of moral responsibility and not just those involving blame (p. 210). It is not immediately obvious just how such a principle is supposed to be formulated. Certainly it will have it that an agent’s being morally responsible for some action A requires both that it was within her control to perform an alternative action B and that B was an action of a certain sort. In cases of blameworthiness, as we have seen, B must have been such that (i) performing B was morally less bad than performing A and (ii) it would have been reasonable for the agent to have considered performing B as an alternative to performing A. But what are to be the correlates to (i) and (ii) in cases of praiseworthiness and responsibility for morally neutral actions? Even if an agent’s being praiseworthy for performing some action A requires that it was within her control to perform an alternative action B that was morally less good than A, do we want to say that it must have been reasonable for her to have considered performing B? Does being morally responsible for performing a morally neutral action, such as deciding to wear one pair of socks rather than another, require that there was an alternative action available that was either morally better or worse? Whatever the case, the important thing to see here is that if moral responsibility requires alternative possibilities of a particular sort, then moral responsibility requires alternative possibilities (simply the). Thus, though there are questions about exactly what sorts of alternative possibility would be included in an unrestricted version of PSA, any unrestricted version of this principle is going to entail PAP.
to be genuinely morally responsible for what he does (such as the ability to properly understand and appreciate moral reasons) is one that proponents of PAP readily endorse. Therefore, it looks as though McKenna fails to offer proponents of PAP any compelling reason to reject this principle in favor of PSA. Indeed, insofar as McKenna is effective in showing PSA to be false, he seems to provide PAP-defenders with a very good reason not to opt for PSA.

### 3.4 Two Kinds of Frankfurt-Style Attack

Although the particular line of argument McKenna advances is, I think, one which can be easily resisted by proponents of PAP, there is another way that limited-blockage cases could be used to challenge the view that moral responsibility requires the ability to do otherwise. Here I make a distinction between two general kinds of Frankfurt-style attack, each of which corresponds to a different way of conceiving of the role of Frankfurt-style examples. In the first kind of Frankfurt-style attack, which I shall call *direct Frankfurt-style attack*, the role of Frankfurt scenarios is to act as counterexamples which falsify PAP. I take it that many (if not most) of those who have used examples of this sort to challenge PAP have (accurately) seen themselves as engaging in this first (direct) kind of attack.\(^{29}\) There is, however, another way that Frankfurt-style cases can be used to challenge PAP. In this second kind of attack, which I shall call *indirect Frankfurt-style attack*, the aim of these cases is not to falsify PAP but to falsify some other principle to which PAP is connected in some important way.

This second (indirect) kind of Frankfurt-style attack is suggested by something John Martin Fischer says. He writes, “Of course, I do not accept the alternative-possibilities control model of moral responsibility. But my contention is that, if you do buy into this traditional picture, then you should also accept that the alternative possibilities must be of a certain sort—they must be sufficiently robust” (1999b, p. 121). According to Fischer, then, one who endorses PAP should also endorse something like the following general principle (call it the Principle of Robust Alternative Possibilities):

PRAP: A person is morally responsible for what he has done only if a robust alternative to it was open to him.

---

\(^{29}\) This is especially, though not exclusively, true of Frankfurt-style arguments offered prior to the line of response developed by Widerker (1995), Kane (1996), and Ginet (1996).
Since PAP does not entail PRAP, proponents of the former are not forced to accept the latter upon pain of contradiction. Why, then, should proponents of PAP endorse PRAP? An earlier passage seems to provide the answer:

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 1999b, p. 99; also see 1994, pp. 140-141; 1999a, pp. 283-284)

In Fischer’s view, it is the intuition that moral responsibility requires a certain sort of control—namely, the type of control that involves possessing not just any alternative possibilities but ones that are distinctly robust—which lies behind the demand for alternative possibilities. It is PRAP which is the fundamental intuition, not PAP. Of course, PRAP entails PAP, so one who accepts the former should also accept the latter; but, again, it is PRAP which is fundamental.

Mounting an indirect Frankfurt-style attack against PAP involves two steps. The first step is to establish a connection between PAP and some other principle, in this case, PSA. This is where Fischer’s point about the basic picture underlying the traditional alternative-possibilities view of moral responsibility comes into play. I suggest that the most plausible way to think about the relationship between PAP and PSA is that PSA grounds PAP in such a way that PAP is plausible only if PSA is plausible. The second step in indirect Frankfurt-style attack is to falsify this other principle (again, in this case, PSA), or at least show that it is implausible. This is where

---

30 In some ways, McKenna appears to echo Fischer’s claim that if moral responsibility requires alternative possibilities it is because it requires alternative possibilities of a certain sort—namely, robust. There is at least one respect, however, in which Fischer and McKenna differ. Whereas Fischer implies that those who accept PAP should also accept something like PRAP (a broadened version of PSA perhaps—one that encompasses all cases of moral responsibility, not just blame), McKenna argues that they ought to opt for PSA instead of PAP, rejecting the latter in favor of the former.
Frankfurt-style cases come in. The role of these cases in indirect Frankfurt-style attack is to act as counterexamples to this other principle to which PAP is related in some important way.

One advantage of indirect Frankfurt-style attack (over against direct Frankfurt-style attack) is that Frankfurt-style cases need not eliminate all alternative possibilities in order to do the work set out for them. All that is required for a case of this sort to count as a genuine counterexample to PSA is that it be one in which an agent has no robust alternatives (of the sort characterized in PSA) to doing that for which she is basically morally responsible. (Like PAP, PSA should be taken to be a thesis about basic moral responsibility.) A further advantage of using Frankfurt-style cases in this way is that any successful flicker-of-freedom-type defense of PSA against Frankfurt-style cases will need to show not only that agents in these cases continue to possess alternative possibilities but that these alternatives are sufficiently robust.

Now, since PAP does not entail PSA, falsifying PSA is insufficient to demonstrate the falsity of PAP. Thus, indirect Frankfurt-style attack is not able to show that PAP is false. Rather than falsifying PAP, the aim of indirect Frankfurt-style attack is to demotivate this principle by showing that the reason(s) for endorsing PAP in the first place ought not to be accepted. This in itself would be significant. To accomplish even this much, however, it must be shown both (i) that PAP is plausible only if PSA is plausible and (ii) that there are possible Frankfurt-style scenarios in which agents without any robust alternative possibilities are nevertheless (basically) morally responsible for what they do. Let us consider these in turn.

3.5 Step One: The Connection between PAP and PSA

Does the plausibility of PAP depend on the plausibility of PSA such that it is reasonable to accept the former only if it is reasonable to accept the latter? For present purposes, I am willing to grant that it is plausible that moral responsibility requires alternative possibilities only if it is also plausible that moral responsibility requires alternatives that are distinctly robust (in the sense that they are able to help ground our attributions of moral responsibility). But must these alternatives be of the sort McKenna identifies in PSA? McKenna holds that there are two general conditions that must be satisfied in order for an alternative to be robust. First, the alternative must be morally significant in such a way that “It would tell us something (different from what we are told in the actual world) about the moral quality of the agent’s conduct were she to have
so acted in this alternative scenario” (2003, p. 204). Second, the alternative must be within the agent’s control.

One way for an alternative possibility to be morally significant is simply for it to have a different moral value—that is, for it to be morally better or worse—than the action an agent actually performs. In this sense, jogging in place, working a crossword puzzle, and humming a tune are all morally significant alternatives to pushing a small child down a flight of stairs or helping an old lady carry her groceries. According to McKenna, however, the mere availability of morally better or worse alternatives, even if they are within an agent’s control, is not itself sufficient for an agent’s possessing robust alternative possibilities. The reason for this, he says, is that “within the very wide spectrum of courses of action with differing moral weights or values, only some in a deliberative context will be relevant to competent moral deliberation and agency” (p. 207). In order to be able to help explain an agent’s being blameworthy, say, for performing some action, it is not enough to point out that there were causally available alternatives that were morally not as bad. It must also be shown that these courses of action were deliberatively significant, which is to say, ones that it would have been reasonable for the agent to have considered as alternatives to what she actually did.

Although McKenna considers it unlikely that one will be able to provide any general principle to distinguish those alternative that are deliberatively significant from those that are not, he offers several cases that he thinks clearly illustrate the distinction. Here is one, which he calls Needed Medication:

In 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. (p. 208)
Even if looking for The Good Stuff in the jar marked ‘aspirin’ (and then finding it and administering it to Daphne) would have been morally better than not doing so, unless it would have been reasonable for Tal, given his agent-relative perspective, to consider this option in his practical deliberation about what to do, McKenna says, it would not have told us anything (different from what we are told in the actual world) about the moral quality of Tal’s will had he pursued this alternative course of action. In order to be morally relevant in a way that would figure into an explanation of an agent’s being morally responsible for what he does (i.e., robust), an alternative must be both morally and deliberatively significant.

I suggest that even if we accept the conditions for robustness laid out above, we should not accept that the plausibility of PAP depends on the plausibility of PSA, at least not in its present form. To see why, it will be helpful to distinguish the following theses, each of which is implicit in PSA:

R1: An alternative possibility is robust only if it is morally significant.
R2: An alternative possibility is robust only if it is deliberatively significant.
R3: An alternative possibility is robust only if it is within the agent’s control.
R4: An alternative possibility is robust only if it is an action.

According to PSA, an agent’s being blameworthy for an action she performs requires not only that she had a morally and deliberatively significant alternative possibility that was within her control but that this alternative was an action. We have just seen the reasons offered for thinking that robust alternatives must be morally significant, deliberatively significant, and within the agent’s control. Grant that R1, R2, and R3 are correct.

What about R4? Why think that only actions can constitute robust alternative possibilities? McKenna never addresses this point, but perhaps the reasoning is something like this: In order for an alternative to be robust, it must be within the control of the agent. Specifically, it must be within an agent’s direct voluntary control, since PSA (like PAP) is meant to be a thesis about basic moral responsibility and agents are basically morally responsible only for that which lies within their direct voluntary control. Given this, perhaps McKenna is thinking that actions (particularly certain mental actions, such as deciding) are the only things over which
agents have direct voluntary control and, so, they alone are plausible candidates for robust alternative possibilities. Fischer appears to have expressed a similar view:

On the traditional alternative-possibilities picture, it is envisaged that an agent has a choice between two (or more) scenarios of a certain sort. In one scenario, he deliberates and forms an intention to perform an act of a certain kind and then carries out this intention in an appropriate way. In at least one other possible scenario, he deliberates and forms an intention to perform a different kind of act (or no act at all) and carries out this intention in an appropriate way. This is what is involved in having robust alternative possibilities, and certainly this is the natural way to think about the sort of alternatives that allegedly ground moral responsibility. (1994, p. 142)

Like McKenna, Fischer maintains that the mere possibility of having something different occur cannot plausibly contribute to an explanation of an agent’s being morally responsible for what she does. To do this, an alternative must be within an agent’s control. Also like McKenna, Fischer seems to presume that an agent will have sufficient control over an alternative only if the alternative is (or at least involves) an action, specifically, a mental act of intention formation (i.e., a decision). Herein lies the problem.

That alternative possibilities can be robust only if they are within an agent’s control is a point well taken. It is certainly plausible that whatever intuitive force accompanies PAP stems from the view that moral responsibility requires what Fischer describes as “the ability to make a difference in the sense of selecting one from various paths the world could take, where these various paths are all genuinely available to the agent” (1999a, p. 283). For unless the availability of various alternatives serves to enhance an agent’s control over the way things go, it is difficult to see how this would be able to help ground attributions of moral responsibility.

What Fischer and McKenna seem to overlook, however, is that agents can see to it that the world takes one path rather than another not only by acting but also by omitting to act. Consider, for instance, a normal case of me sitting at the bus stop, waiting for my bus to arrive. Once it arrives there are various paths the world could take that are genuinely available to me. One path involves me getting on the bus and making it home in time for dinner; and I can see to it that the world takes this path by deciding to get on the bus, which it is within my direct
voluntary control to do. There are also a number of other paths that do not involve me getting on the bus or making it home in time for dinner; and I can see to it that the world takes one of these paths rather than the first one. One way I can do this is by making a different decision—either a decision not to get on the bus or a decision to do something else that is incompatible with my getting on the bus then. Importantly, however, this is not the only way that I can control which path the world takes. Another way I can see to it that the world takes a path that does not involve me getting on the bus is by simply omitting 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.

Given that agents can have at least as much control over their omissions as they can over their actions, there appears to be no good reason for insisting that only actions are capable of constituting robust alternative possibilities. To the extent, then, that PSA presumes that omissions cannot be robust alternatives, we should not accept that the plausibility of PAP is in any way linked to the plausibility of PSA. This problem can be avoided by revising PSA as follows:

\[
\text{PSA*: An agent } S \text{ is morally blameworthy for performing action } A \text{ at } t \text{ only if she had within her control at } t \text{ not performing } A \text{ (either by performing an alternative action } B \text{ or by simply omitting to perform } A \text{) such that (1) not performing } A \text{ at } t \text{ was morally less bad than performing } A \text{ at } t, \text{ and (2) it would have been reasonable for } S \text{ to have considered not performing } A \text{ at } t \text{ as an alternative to performing } A \text{ at } t \text{ given } S\text{'s agent-relative deliberative circumstances.}
\]

Grant, for the sake of argument, that PAP is plausible only if PSA* is also plausible. As before, this will be useful in demotivating PAP only if there are cases that show PSA* to be false (or at least implausible). Let us now consider whether Brain Malfunction is a case of this sort.

### 3.6 Step Two: PSA* and Brain Malfunction

Assuming that Casper is (basically) morally responsible for deciding to press the Bad button, Brain Malfunction will be a successful counterexample to PSA* only if there were no
alternatives to deciding to press the Bad button that, in addition to being within his control, were both morally and deliberatively significant. On the face of it, it certainly looks as though Casper does have alternatives of just this sort. For one thing, Casper could have decided that, whatever else he did, he most definitely would not press the Bad button. Pretty plainly, deciding not to press the Bad button would have been morally less bad than deciding to press the Bad button; and had he done this, which it was within his control to do, Casper would have avoided (being blameworthy for) stealing money from his peers. According to McKenna, however, deciding not to press the Bad button was not a deliberatively significant alternative:

How can it be shown that Casper’s simply deciding not to press the Bad button was deliberatively insignificant? 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. (2003, p. 211)

The problem with this attempt to rule out deciding not to press the Bad button as a deliberatively significant alternative is that it relies on an understanding of the reasonableness condition in PSA(*) that McKenna himself explicitly rejects:

What counts as reasonable in the second condition of PSA is left vague. It is correspondingly vague in our moral practices. Hence, the vagueness is perfectly acceptable. But one particular manner in which the vagueness might appear problematic turns on a possible misreading of PSA according to which the scope of reasonableness from an agent-relative perspective is limited to the values and standards endorsed by the agent. What the white supremacist finds reasonable, for instance, differs from what one
ought to find reasonable. But the objection turns upon the mistaken assumption that agent-relative perspectives are not subject to objective criteria of rationality and truth. (p. 209)

McKenna attempts to establish that deciding not to press the Bad button was not a deliberatively significant alternative—that is, that it would not have been reasonable for Casper to have considered deciding not to press the Bad button as an alternative to deciding to press it, given his agent-relative deliberative circumstances—simply by stipulating that Casper would not have considered this to be a relevant option. But as McKenna himself makes quite clear, an alternative’s being deliberatively significant is a function of what an agent ought to find reasonable (or relevant), not what an agent in fact finds reasonable (or relevant). This is as it should be.

The whole reason for restricting the range of morally significant alternatives relevant to assessing agents’ moral responsibility in the first place, according to McKenna, stems from the idea that it is unreasonable to expect competent moral agents to be aware of all morally significant options. “It cannot be a demand upon moral agents,” he says, “that they possess the epistemic powers of deities or the wisdom of saints” (p. 208). Fair enough. Surely, however, one need not possess the wisdom of Solomon, let alone divine omniscience, to see that deciding not to steal from one’s peers is a morally significant alternative to deciding to steal from them. Unlike looking for The Good Stuff in a jar marked ‘aspirin’, deciding 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.31

31 Perhaps what McKenna means to suggest is not just that Casper would not have found this to be a reasonable alternative but that he would have been correct: Given that Casper is presented
A different tack, which McKenna mentions but does not pursue, would be to modify the case so that the alternative of deciding not to press the Bad button is also blocked by a brain lesion (p. 211). This would rule out deciding not to press the Bad button as a robust alternative possibility, not by rendering it deliberatively insignificant, but by removing it from the scope of Casper’s voluntary control. Even then, however, Casper would continue to possess robust alternatives. Since there is nothing that forces, nor is there anything that stands ready to force, Casper to decide to press the Bad button, it is within Casper’s control to simply omit to decide to do so. Now, given that in this modified version of Brain Malfunction 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. But 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.

The kind of omission I here have in mind corresponds to a class of action that some philosophers have dubbed “nonintentional action” (Mele and Moser 1994; Mele and Sverdlik

with the option of pressing the Good button, it would in fact have been unreasonable of him to have considered simply deciding not to press the Bad button and yet not press the Good button. In that case, it should be emphasized that, as far as PSA(*) is concerned, what matters here is not whether it would have been reasonable for Casper to have considered simply deciding not to press the Bad button as an alternative to deciding to press the Good button but, rather, whether it would have been reasonable for Casper to have considered this as an alternative to deciding to press the Bad button. Even if it would not have been reasonable for Casper to have considered simply deciding not to press the Bad button as an alternative to deciding to press the Good button, it unquestionably would have been reasonable for him to have considered this as an alternative to what he actually did.

32 Some theorists have defended a “simple view” of intentional action according to which an agent intentionally performs an action A only if she intends to A (Adams 1986; McCann 1986; 1989; 1991). I believe others have shown that there are good reasons for rejecting this view (Bratman 1987; Mele 1989; 1992). Though an A-ing’s being intentional requires that it be suitably related to some sort of intention (e.g., an intention to try to A), it does not require an intention to A. Plausibly, the same is true of intentional omission (see Clarke 2010). In saying that an agent who is prevented from forming an intention to omit to A may be unable to intentionally omit to A, I do not mean to suggest that intentionally omitting to A requires intending to omit to A (what one might think of as a “simple view” of intentional omission). But since any intention in virtue of which Casper’s omitting to decide to press the Bad button might be intentional could be blocked by further, fairly trivial modifications to the case, this is not an issue I intend to make any hay of here.
1996). To illustrate, consider a runner who runs a marathon and, in so doing, wears down the soles of her running shoes. Suppose that, as she is running, she is consciously aware of the fact that she is wearing down the soles of her shoes, even though she has no intention of doing so (nor does she have any other intention whose content includes the wearing down of the soles of her shoes, e.g., an intention to try to wear down the soles of her shoes). Because the runner is both in control of her actions and aware that she is wearing down the soles of her shoes as she is doing it, it would seem to be a mistake to say that she does this unintentionally. At the same time, however, it would also be inaccurate to say that she intentionally wears down the soles of her shoes since 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. Fortunately, as some theorists point out, there is a middle ground: “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” (Mele and Sverdlik 1996, p. 274; also see Mele and Moser 1994, p. 45). I suggest that, like actions, omissions too can be nonintentional in this way. Just as one can knowingly and voluntarily perform an action and yet not perform it intentionally, one can knowingly and voluntarily omit to perform an action without intentionally omitting to perform it.34

---

33 This example comes from Michael Bratman (1987, p. 123).

34 Here it is worth noting that typical examples of nonintentional action mostly involve side-effects of actions performed intentionally: a runner who wears down the soles of her shoes in the course of running a marathon; a sniper who alerts the enemy to his presence in the course of shooting at his target (Harman 1976, p. 433); a dentist who inflicts pain upon her patient in the course of performing a needed medical procedure (Mele and Sverdlik 1996, p. 274). As such, they are only indirectly within an agent’s control. Though these nonintentional actions are indeed performed voluntarily, their voluntariness derives from their relation to other, intentional actions, which are within the agent’s direct voluntary control. By contrast, 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. (This is not to say that omitting to A entails knowingly and voluntarily omitting to A. Certainly there are other conditions that must be met in order for an agent who is omitting to A to be knowingly and voluntarily omitting to A—namely, (i) that the agent is aware he is omitting to A and (ii) that omitting to A is within his voluntary control. The point here is that there is nothing else that he needs to do in order for his omitting to A to count as a case of knowingly and voluntarily omitting to A.) This potentially suggests the following asymmetry between nonintentional actions and omissions: Whereas nonintentional actions are only indirectly within...
With this distinction in mind, it becomes clear that Casper’s being blameworthy for deciding to press the Bad button is entirely in keeping with the conditions for blameworthiness set out in PSA*: 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. Thus, even when modified to prevent Casper from deciding not to press the Bad button (or in any other way intentionally omitting to do so), Brain Malfunction fails to demonstrate that blameworthiness (let alone moral responsibility) does not require robust alternative possibilities.

35 No doubt McKenna would maintain that, just as with deciding not to press the Bad button, one could describe the case in such a way that simply omitting to decide to press the Bad button was not a deliberatively significant alternative—namely, by stipulating that Casper would not have considered this to be a relevant option. But, just as with deciding not to press the Bad button, the problem with this response is that, even according to McKenna’s own view, an alternative’s being deliberatively significant is a function of what an agent ought to find reasonable (or relevant), not what an agent in fact finds reasonable (or relevant). It is always reasonable for an agent to consider not doing that which she knows to be wrong as an alternative to doing it.

36 Some readers may observe a general resemblance between the criticism of the limited-blockage strategy developed in this chapter and Carl Ginet’s (2002) response to a Frankfurt-style case offered by Derk Pereboom (2001) called “Tax Evasion.” Both maintain that agents in these cases had robust alternative possibilities having to do with their ability to have omitted to act as they did. There are at least two ways, however, in which my argument differs significantly from Ginet’s. First, Ginet’s response pushes the so-called “timing worry” that faces a number of Frankfurt-style cases. He argues that, although the agent in Tax Evasion could not have avoided deciding to cheat on his taxes, it was possible for the agent to have refrained from making that decision at the time that he did (and so been forced to make the decision at a slightly later time). Responses of this sort apply only to cases in which there are ensuring conditions in place to guarantee that the agent cannot avoid making the decision he actually makes. Thus, it cannot be applied to limited-blockage cases (such as Brain Malfunction), which are meant to allow agents a wide range of (nonrobust) alternatives to deciding and acting as they actually did (let alone the time at which they made the decisions). Second, and more important, Ginet’s response accepts Pereboom’s assumption that a robust alternative must be (or at least involve) an intentional action (viz., a “willing”). He argues that the agent in Tax Evasion could have refrained from deciding at $t_1$ to cheat on the grounds that the agent could have done something else (brought
There is one final move that remains open to one who wishes to attempt to preserve Brain Malfunction as a counterexample to PSA*, which is to further modify the case in a way that eliminates Casper’s ability to knowingly and voluntarily omit to decide to press the Bad button. The only way to eliminate the possibility of Casper knowingly and voluntarily omitting to decide to press the Bad button (aside from rendering him unconscious or otherwise not in control of his behavior) is to guarantee that he decides to press it. And the only way to accomplish that without undermining the claim that Casper is morally responsible for what he does (assuming this is possible) is to block his access to all other actional pathways and include an ensuring mechanism that will force Casper to decide to press the Bad button unless he does so on his own. At this point, however, Brain Malfunction ceases to be a limited-blockage example (or, at least, if it is, its being so in no way distinguishes it from various other kinds of Frankfurt-style example to which it is meant to be an alternative).\(^{37}\)

### 3.7 Conclusion

McKenna maintains that even if there is some weakness in Brain Malfunction that prevents it from being a successful counterexample to PSA(*), “It is difficult to see what theoretical basis there could be for denying that [some] Frankfurt example could be constructed that closed off all and only the deliberatively significant alternatives while leaving open some of the insignificant ones” (2003, p. 212). Perhaps the reason for McKenna’s confidence in the eventual success of some limited-blockage case, even if not Brain Malfunction, stems from the implicit assumption that only actions can constitute robust alternative possibilities. But once we see that omissions too are capable of playing this role, it becomes clear that there are reasons for thinking that the

---

37 Whether Brain Malfunction, so modified, would still qualify as a limited-blockage example depends on whether it is an essential feature of cases of this sort that agents have access to a wide range of alternative possibilities (“oodles and oodles” of them, as McKenna puts it). If not—if the sole defining characteristic of these cases is that they eliminate all and only robust alternatives—then, while such a version of Brain Malfunction would technically count as a limited-blockage example, it would be indistinguishable in effect from other modified Frankfurt-style cases (just as, when dealing with a barrel of all red apples, a policy of removing all the apples and a policy of removing all and only the red apples both result in an empty barrel).
limited-blockage strategy itself is theoretically incapable of doing the work set out for it. Omitting to do that which one knows to be wrong is always a deliberatively significant alternative: 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 doing that which she knows to be wrong. Furthermore, omissions can be within agents’ direct voluntary control even when they are not intentional. Thus, so long as it remains open to agents in these cases to knowingly and voluntarily omit to decide to act as they do, they will continue to possess robust alternative possibilities. The only way to close off this robust (and so deliberatively significant) alternative—namely, by guaranteeing that these agents will decide to act as they do—requires closing off all the alternatives within the agent’s voluntary control, including the deliberatively insignificant ones. The virtue of the limited-blockage strategy is supposed to be that “limited blockage cases openly grant libertarian freedom involving alternative possibilities” and, “therefore, clearly do not require any special considerations that might tacitly import a deterministic relation between agent and action” (p. 212). Once limited-blockage cases are modified to eliminate the possibility of agents knowingly and voluntarily omitting to decide to act as they do, this advantage is lost.
CHAPTER FOUR

ROBUST FLICKERS OF FREEDOM

4.1 Introduction

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. The greatest challenge currently facing this seemingly intuitive principle comes from a class of cases developed by Harry Frankfurt (1969), commonly known as Frankfurt-style cases. The original intent in offering these cases was to falsify PAP by providing possible scenarios in which agents appear to be morally responsible for performing some action that, owing to the special circumstances in which they find themselves, they were unable to avoid performing. One way to defend PAP against falsification by cases of this sort is to call into question the claim that agents in these cases are genuinely morally responsible for what they do (at least with respect to those actions whose performance is unavoidable for them). Unless the actions agents could not have avoided performing are also ones for which they are morally responsible, these cases will fail to falsify the view that the ability to do otherwise is a necessary condition for moral responsibility.

Alternatively, one could grant that agents are morally responsible for what they do in these cases but deny that they could not have done otherwise. According to this line of argument, even though Frankfurt-style cases manage to significantly restrict agents’ ability to do otherwise, they do not eliminate it in the way that falsifying PAP would require. This latter strategy has come to be known as the flicker defense of alternative possibilities.38

In Chapter 2 I defended a version of the flicker strategy that focuses on the fact that agents in these cases do what they do on their own, just as they would have had they been in ordinary (non-Frankfurtian) circumstances, which is something regarding which they could have done otherwise. Even though it is not possible for these agents not to perform the actions they

38 The term ‘flickers of freedom’ was coined by John Martin Fischer (1994, pp. 137-47) to refer to those alternatives that remain open to agents in Frankfurt-style cases.
do, since it remains up to them whether to perform these actions on their own, the fact that they are morally responsible for them poses no threat to PAP.

Critics of the flicker strategy have argued that the alternative possibilities that remain open to agents in Frankfurt-style cases—including the ability to omit to perform certain actions on their own—are not adequate to the task of rebuffing Frankfurt-style attack. In order to neutralize the threat facing PAP, it is not enough, they maintain, for agents in these cases to have alternative possibilities—the alternatives must be robust. That is, they must be morally significant in a way that helps to account for the agents’ moral responsibility for what they actually do. Insofar, then, as Frankfurt-style cases are able to restrict agents’ ability to do otherwise to alternatives that are insufficiently robust in this sense, they will be successful in undermining PAP (assuming these agents are morally responsible for what they do). Eliminating all alternative possibilities, including those that are not robust, they argue, is not required.

My aim in this chapter is to consider whether there is any good reason for thinking both (i) that the success of the flicker strategy requires appealing to distinctly robust alternative possibilities and (ii) that there are (or at least can be) Frankfurt-style cases that eliminate all such alternatives (again, without undermining agents’ moral responsibility for what they do). I shall argue that whereas demonstrating that Frankfurt-style cases fail to constitute genuine counterexamples to PAP does not require showing that at least some of the alternatives open to agents in these cases are distinctly robust, there is another kind of Frankfurt-style attack where the demand for robustness is appropriate. There the goal in offering Frankfurt-style cases is not to falsify PAP (perhaps because this is viewed as a less promising venture) but, rather, to undercut its motivation by challenging certain intuitions taken to be the grounds for this principle. As I shall further argue, however, even if all that is needed to demotivate (though not falsify) PAP in this way are Frankfurt-style scenarios that manage to eliminate all robust alternative possibilities, there are no (nor can there be) cases of this sort. So long as it remains up to agents in these cases whether to on their own do the things for which they are morally responsible, they will continue to possess robust alternative possibilities. And since having it be up to the agents whether they perform the relevant actions on their own is a crucial feature of Frankfurt-style cases—one which cannot be eliminated without undermining the judgment that they are genuinely morally responsible for what they do—the demand for robust flickers of freedom is one that this particular version of the flicker strategy will always be able to meet.
4.2 Frankfurt-Style Cases and Flickers of Freedom

A genuine counterexample to PAP will be a case in which an agent is morally responsible for performing some action she could not have done otherwise than perform. Because PAP is generally understood to be a thesis about basic (i.e., direct, or non-derivative) moral responsibility, the relevant actions in these cases tend to be mental acts of deciding, which are commonly taken to be the things for which we are basically morally responsible and from which our responsibility for overt physical actions (and their consequences) derives. To ensure that agents cannot do otherwise than decide to act as they actually do, typical Frankfurt-style cases include the presence of a counterfactual intervener, someone who stands ready and able to force the agent to decide to act in the desired way in the event that the agent is not going to do so on her own. Crucially, however, the counterfactual intervener’s presence is unknown to the agent and plays no role in the agent’s actual behavior, lest it undermine the intuition that the agent is morally responsible for what she does.

The tricky part for cases of this sort comes in identifying the precise conditions under which the counterfactual intervener would intervene. If the counterfactual intervener has to wait to see how the agent actually decides, then the agent retains the ability to do otherwise, and PAP is unthreatened. In order to guarantee that the agent will decide to act in the desired way without actually compelling the agent’s behavior, the counterfactual intervener must have some way of knowing in advance what the agent will do if left to act on her own. Toward this end, traditional Frankfurt-style cases posit the existence of a prior sign—something beyond the agent’s voluntary control (such as a blush or a cluster of neural firings) that indicates how the agent would freely act, were no intervention to occur.

Counterexamples to PAP need not be restricted to cases involving actions. Since PAP is generally taken to apply to both actions and omissions, a case in which an agent is genuinely morally responsible for omitting to perform some action she could not have done otherwise than omit to perform would also serve as a counterexample to this principle. Here I focus on cases involving moral responsibility for actions (as opposed to omissions) merely for the sake of simplicity.

An agent is basically morally responsible for some action she performs when her responsibility for that action is not inherited from (or in virtue of) her responsibility for other actions. Moral responsibility that is inherited in this way is derivative, or indirect (see Mele 2006, p. 86). For arguments that agents can be basically morally responsible for things other than certain mental actions, see Adams 1985 and Smith 2005, 2006.
The trouble with including indicators of this sort is that they are capable of doing the work set out for them in traditional Frankfurt-style cases only by presupposing a deterministic relation between an agent’s free action and its causal antecedents—a relation incompatibilists flatly reject (see Widerker 1995; Kane 1996; Ginet 1996). This problem can be seen as posing the following dilemma. Either the occurrences of these prior signs (or that which they indicate) make it merely probable that agents in these cases will decide to act in the desired ways, or they deterministically lead to the agents so deciding. If the occurrences of these signs make it merely probable that the agents will decide to act in the desired ways, then it remains causally open to them not to decide to act in this way after all. Insofar as this is the case, however, the agents will continue to possess alternative possibilities with respect to their actions, and so their bearing moral responsibility for these actions is perfectly compatible with PAP. If, on the other hand, the occurrences of these signs (or that which they indicate) will deterministically lead to the agents deciding to act in the desired ways, then it will no longer remain causally open to the agents not to decide to act as they do. But since having one’s actions causally determined precludes one from being morally responsible for them, according to incompatibilists, that agents in these cases are nevertheless morally responsible for what they do cannot be asserted without begging any questions.  

Either way, then, traditional Frankfurt-style scenarios fail to constitute genuine counterexamples to PAP. 

This particular line of objection has prompted a number of attempts to modify traditional Frankfurt-style cases in ways that will enable them to do the work set out for them by Frankfurt without employing the use of (determinative) prior signs. Here is a modified Frankfurt-style case offered by Alfred Mele and David Robb: 

---

41 For arguments to the effect that there is in fact nothing illicit in the way traditional Frankfurt-type counterexamples make use of prior signs, see Fischer 1999; Haji and McKenna 2004.

42 It should be noted that, relying as it does on the decidedly incompatibilist intuition that determinism and moral responsibility are mutually exclusive, this is not a line of argument available to compatibilist defenders of PAP. Compatibilists who wish to resist the conclusion of Frankfurt’s argument cannot do so on the grounds that it is question-begging to assert that agents can be morally responsible for their actions despite their being causally determined.

43 See, for example, Haji 1998; Hunt 2000; McKenna 2003; Mele and Robb 1998, 2003; Pereboom 2000; Stump 1996.
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. (1998, pp. 101-2; footnotes omitted)

Rather than rely on the presence of some prior sign concerning what Bob will freely do if left to act on his own, Mele and Robb employ a particular kind of causal preemption to guarantee that an intervention will take place if and only if Bob does not decide on his own to steal Ann’s car (and is capable of making a decision at that time). The significance of the preemption’s occurring precisely at $t_2$ (the moment of action) is that it ensures that at no point leading up to and including $t_2$ is it possible for Bob to do otherwise than decide at $t_2$ to steal Ann’s car. Because the process Black initiates in Bob’s brain is insensitive to anything Bob does prior to $t_2$, Mele and Robb avoid the need to controversially presuppose a deterministic relation between Bob’s action and any of its causal antecedents. And because the process initiated by Black is causally preempted by Bob at $t_2$, and so plays no causal role in Bob’s actual decision to steal Ann’s car, there appears to be no reason to deny that Bob can be genuinely morally responsible for this decision.

In Chapter Two I argued at length that even if the modified Frankfurt-style case just described is successful in avoiding the dilemma facing traditional prior-sign cases, it nevertheless fails to falsify PAP. I will not rehearse all that I said there, but the core of my argument can be put as follows. It is generally accepted by those on both sides of this debate that PAP is properly understood as a thesis about basic (i.e., direct, or nonderivative) moral
responsibility. Indeed, it is only when PAP is understood in this way that the “need” for
Frankfurt-style examples arises in the first place. Were it taken to apply to instances of nonbasic
(i.e., indirect, or derivative) moral responsibility as well, falsifying this principle would not
require so much talk of nefarious neuroscientists and their science-fictional devices. Standard
drunk-driver cases would do the trick. So understood, then, PAP claims that a person is
basically morally responsible for what he has done only if he could have done otherwise.
Importantly, it is the referent of the former which fixes that of the latter and not the other way
around. That is to say, that than which a person must have been able to do otherwise, according
to PAP, is determined by that (whatever it is) for which he is basically morally responsible.

In order for a Frankfurt-style case to count as a genuine counterexample to PAP, it must
be one in which an agent could not have done otherwise than that for which he is basically
morally responsible. The example provided by Mele and Robb is not such a case. Even
supposing it is true both that Bob could not have done otherwise than decide to steal Ann’s car
and that the deterministic process initiated by Black played no causal role in bringing about
Bob’s decision, it is not deciding to steal Ann’s car for which Bob is basically morally
responsible. This is not to deny, I hasten to add, that Bob is genuinely morally responsible for
deciding to steal Ann’s car, only that he is basically morally responsible for this. It is deciding on
his own to steal Ann’s car for which Bob is basically morally responsible and from which his
moral responsibility for so deciding simpliciter derives.

One reason for thinking that Bob is basically morally responsible not for deciding
simpliciter to steal Ann’s car but for so deciding on his own is implicit in Mele and Robb’s case
itself. Like virtually all other Frankfurt-style scenarios (both in their traditional and modified
forms), Mele and Robb’s case focuses on a mental act of deciding rather than an overt physical
action. This is no accident. Instead, it is the product of two widely shared assumptions. The first

44 See Mele 2006, pp.84-86

45 Some philosophers have argued that the particular modified Frankfurt-style scenario offered
by Mele and Robb, insofar as it involves actually blocking off every alternative to Bob’s
deciding to steal Ann’s car, is one in which the agent’s action appears to be causally determined
(Kane 2000; Pereboom 2001; Ekstrom 2002). Others have challenged the coherence of the
occurrent preemption on which their case relies (Widerker 2000). Here I set these worries aside
in order to show that, even if we grant that these worries can be allayed, PAP is able to
accommodate cases of this sort. For their reply to these concerns, as well as a detailed account of
the type of mechanism they have in mind, see Mele and Robb 2003.
is that PAP is properly understood as a thesis about basic moral responsibility. The second is that responsibility for overt physical actions (such as raising one’s arm) derives from responsibility for certain mental actions (such as deciding to raise one’s arm) much in the same way that responsibility for the consequences of one’s actions derives from one’s responsibility for the actions themselves. On this view, if one is morally responsible both for deciding to raise one’s arm and for raising one’s arm, it is the former for which one is basically morally responsible, whereas one is only derivatively responsible for the latter. This variation in the kind of responsibility agents bear for certain mental actions as compared to overt physical actions is best explained by a difference in control. The successful performance of even simple overt actions like raising one’s arm is much more vulnerable to environmental impediments than that of simple mental actions like deciding to raise one’s arm. It is because of the increased control agents have over certain mental actions, such as deciding, that it is these, rather than overt physical actions, for which they are basically morally responsible.

Things are a bit different in Frankfurt-style scenarios. Looking at Mele and Robb’s case, we can see that Bob has the same degree of control over whether he will steal Ann’s car as he does over whether he will decide to steal Ann’s car—namely, none. That Bob will decide to steal Ann’s car and that he will steal it have been guaranteed, and there is nothing he can do to prevent either of these things from occurring. Even in this case, however, whether Bob will decide on his own to steal Ann’s car, and whether he will steal Ann’s car on his own, remains entirely up to him. Thus, the same reasons typically taken to support the view, implicit in Frankfurt-style examples, that one’s responsibility for performing some overt action A derives from one’s responsibility for performing some mental action, such as deciding to A, equally suggest that, at least in Frankfurt-style cases, one’s responsibility for deciding to A derives from one’s responsibility for so deciding on one’s own, for which one is basically morally responsible.

46 Strictly speaking, Mele and Robb’s case ensures that at \( t_2 \) Bob will decide to steal (and then steal) Ann’s car provided that Bob is not generally incapable of making a decision at that time (say, because he is unconscious or dead). Thus, since in their description of the case there is no mechanism in place to prevent Bob from rendering himself incapable of making any decision at \( t_2 \), up until \( t_2 \) it remains causally open to Bob to render himself incapable of making any decision at that time, including a decision to steal Ann’s car. It appears, however, that it would be easy enough to describe the case in a way that rules out this possibility as well.

47 I leave it open whether agents’ moral responsibility for A-ing is always derivative from their responsibility for A-ing on their own or whether agents are sometimes basically morally responsible.
So, even if Bob is genuinely morally responsible for deciding to steal Ann’s car, his responsibility for this derives from his responsibility for deciding on his own to steal Ann’s car. Of course, since deciding on his own to steal Ann’s car is something it was within Bob’s power not to have done, the judgment that he is (basically) morally responsible for doing this is consistent with the truth of PAP.\footnote{I take it to be a virtue of this particular version of the flicker strategy over against other versions that it does not require reformulating PAP in a way that is likely to spur complaints of gerrymandering (see Shabo 2007, p. 69; cf. Timpe 2006) or diminish its intuitive force (by making it overly complex). Nor does it commit its proponents to maintaining either that A-ing and A-ing on one’s own are distinct actions and events (cf. van Inwagen 1978) or that agents in Frankfurt-style cases are morally responsible only for A-ing on their own and not at all for A-ing \textit{simpliciter} (cf. Naylor 1984).}

It is important to note that none of the foregoing hinges upon anything peculiar to Mele and Robb’s particular example. Rather, it capitalizes on two features that are essential to any Frankfurt-style scenario: The agent acts on his own, and it was possible for the agent not to have done so. That agents in these cases act on their own, just as they would have had no ensuring mechanism been in place, is what is supposed to drive the intuition that they are morally responsible for what they do, without which there is no threat to PAP. Furthermore, omitting to perform the relevant action on one’s own is a genuine alternative possibility in all Frankfurt-style scenarios (and one which in modified cases lies safely within the scope of the agents’ direct voluntary control). Indeed, it is precisely because it is possible for agents in these cases not to perform the relevant actions on their own that the ensuring mechanisms are there to begin with; otherwise they would be superfluous. Since this alternative possibility is ineliminable, the particular version of the flicker strategy employed here can be brought to bear against any modified Frankfurt-style example. Insofar as it is successful, the upshot is not only that PAP has yet to be falsified by any of the Frankfurt-style cases currently on offer but that cases of this kind are theoretically incapable of falsifying PAP.
4.3 Why Must Flickers Be Robust?

Recall that, according to Fischer, neutralizing the threat to PAP from Frankfurt-style cases requires not only that agents in these cases could have done otherwise than that for which they are (basically) morally responsible but that the existence of these alternatives is able to help account for why agents would, in virtue of possessing these alternatives, be morally responsible for what they actually do. As he puts it, the alternatives must be sufficiently robust.\(^{49}\) According to Fischer, one who endorses PAP should also endorse something like the following general principle (call it the Principle of Robust Alternative Possibilities):

\[
\text{PRAP: A person is morally responsible for what he has done only if he could have done otherwise and his doing otherwise was a robust alternative possibility.}
\]

In Fischer’s view, it is the intuition that moral responsibility requires a certain sort of control—namely, the type of control that involves possessing not just any alternative possibilities but ones that are distinctly robust—which lies behind the demand for alternative possibilities. It is PRAP which is the fundamental intuition, not PAP. Of course, PRAP entails PAP, so one who accepts the former should also accept the latter; but, again, it is PRAP which is fundamental.

In the previous chapter I explained how Fischer’s point could be used as part of an indirect Frankfurt-style attack on PAP.\(^{50}\) Recall that mounting an indirect Frankfurt-style attack against PAP involves two steps. The first step is to establish a connection between PAP and some other principle, in this case PRAP. This is where Fischer’s point about the basic picture underlying the traditional alternative-possibilities view of moral responsibility comes into play. I suggest that the most plausible way to think about what Fischer says concerning the relationship between PAP and robustness is that PRAP (or some other similar principle) grounds PAP in such a way that PAP is plausible only if PRAP is plausible. The second step in indirect Frankfurt-style attack is to falsify PRAP (or at least show that it is implausible). This is where Frankfurt-style cases come in. The role of these cases in indirect Frankfurt-style attack is to act as counterexamples to PRAP.

\(^{49}\) See section 2.7.
\(^{50}\) See section 3.4.
For present purposes, I am willing to grant that PAP is plausible only if PRAP is also plausible and that falsifying PRAP would have the result of demotivating PAP in the way suggested above. In what follows, I want to consider whether there could be Frankfurt-style scenarios in which agents without any robust alternatives are nevertheless (basically) morally responsible for what they do. First, however, we must get clearer on the nature of robustness.

4.4 What Does Robustness Require?
The robustness challenge, originally offered by Fischer, has been echoed by a number of prominent Frankfurt-defenders to address the flicker strategy. There is general agreement that a robust alternative is one which is capable of grounding ascriptions of moral responsibility—that is, one which is relevant in and of itself to explaining why an agent is morally responsible for what she has actually done. There is some variation, however, concerning what is required for an agent to have alternatives that are robust in this sense. Whereas Fischer focuses on the requirement that robust alternatives be within agents’ (voluntary) control, others emphasize the importance of satisfying certain epistemic conditions. Michael McKenna maintains that, in addition to being within an agent’s voluntary control, a robust alternative must be such that “It would tell us something (different from what we are told in the actual world) about the moral quality of the agent’s conduct were she to have so acted in this alternative scenario” (2003, p. 204). To do this, he says, an alternative must be both morally significant (i.e., morally better or worse than the action the agent actually performs) and deliberatively significant (i.e., reasonable to expect the agent to have considered as an alternative to what she actually does, given her agent-relative perspective). According to Derk Pereboom, an agent’s having a robust alternative requires that the agent “could have willed something other than what she actually willed such that by willing it she would thereby have been precluded from the moral responsibility she actually has for the action” (2003, p. 188).

For present purposes, I want to grant that robustness requires everything that Fischer, McKenna, and Pereboom say it does—with one exception. They all appear to presuppose that a robust alternative must (at least partly) consist in an action. Why think that only actions can constitute robust alternative possibilities? None of them explicitly addresses this point, but perhaps the reasoning is something like this: In order for an alternative to be robust, it must be within the control of the agent. Specifically, it must be within an agent’s direct voluntary control,
since PRAP (like PAP) is meant to be a thesis about basic moral responsibility and agents are basically morally responsible only for that which lies within their direct voluntary control.

Therefore, since actions (particularly certain mental actions, such as deciding) are the only things over which agents have direct voluntary control, they alone are plausible candidates for robust alternative possibilities. Like Fischer, McKenna and Pereboom maintain that the mere possibility of having something different occur cannot plausibly contribute to an explanation of an agent’s being morally responsible for what she does. To do this, an alternative must be within an agent’s control. Also like Fischer, they seem to presume that an agent will have sufficient control over an alternative only if the alternative is (or at least involves) an action, specifically, a mental act of intention formation (i.e., a decision).⁵¹ Herein lies the problem.

That alternatives can be robust only if they are within an agent’s control is a point well taken. It is certainly plausible that whatever intuitive force accompanies PAP stems from the view that moral responsibility requires what Fischer describes as “the ability to make a difference in the sense of selecting one from various paths the world could take, where these various paths are all genuinely available to the agent” (1999a, p. 283). Unless the availability of various alternatives enhances an agent’s control over the way things go, it is unclear how this would help ground attributions of moral responsibility. As I argued in the previous chapter, however, agents can see to it that the world takes one path rather than another not only by acting but also by omitting to act. Omissions too can be within agents’ direct voluntary control.⁵² Given that agents can have at least as much control over their omissions as they can over their actions, there appears to be no justification for insisting that only actions are capable of constituting robust alternatives. Aside from that, let us assume that robustness requires everything else Fischer, McKenna, and Pereboom say it does. Again, this will be helpful in demotivating PAP only if it can be shown that robust alternatives of this sort are unnecessary for moral responsibility.

---

⁵¹ Though Fischer seems to allow that an agent’s performing no act at all could qualify as a robust alternative, this is conditioned on its being the result of a decision to do so (1994, p. 142; quoted above). Thus, he grants that robust alternatives can at least partly consist in omissions but only insofar as they also involve an act of intention formation (i.e., a decision).

⁵² See pp. 42-44.
4.5 Robust Flickers of Freedom

Here we come to the central task of this chapter, which is to show that indeterministic Frankfurt-style cases are theoretically incapable of falsifying PRAP.\(^ {53} \) To begin, assume that, like PAP,

\( (1) \) PRAP is a thesis about basic moral responsibility.

Also like PAP, PRAP is a conditional in which the referent of the consequent is fixed by that of the antecedent (and not the other way around). That is, PRAP requires that a person had a robust alternative to doing *that for which he is basically morally responsible*. Consequently,

\( (2) \) A counterexample to PRAP must be a case in which a person had no robust alternative to doing *that* for which he is basically morally responsible.

The first major step in this argument is to show that

\( (3) \) Agents in Frankfurt-style cases are basically morally responsible for deciding on their own to act as they do (not for so deciding *simpliciter*).

Earlier I argued that the same reasons that support the view that moral responsibility for overt actions (such as raising one’s arm) derives from responsibility for certain mental actions (such as deciding to raise one’s arm) equally support the view that, at least in Frankfurt-style cases, agents’ responsibility for deciding to perform the relevant actions derives from their responsibility for so deciding on their own (for which they are basically morally responsible).

There is another reason for thinking that agents in Frankfurt-style cases are basically morally responsible for deciding on their own to act as they do, not for so deciding *simpliciter*, which we can see by considering the following question. Assuming that Bob (from Mele and Robb’s case) is morally responsible both for deciding on his own to steal Ann’s car and for deciding to steal Ann’s car, what should we think about the relationship between his

\(^ {53} \) Deterministic Frankfurt-style cases are unable to serve as good counterexamples to PRAP for the same reason they are unable to do so for PAP (see pp. 4-5).
responsibility for the former and his responsibility for the latter? There seem to be three possibilities:

(i) Bob’s moral responsibility for deciding to steal Ann’s car derives from his responsibility for deciding on his own to steal Ann’s car.

(ii) Bob’s moral responsibility for deciding on his own to steal Ann’s car derives from his responsibility for deciding to steal Ann’s car.

(iii) Neither Bob’s moral responsibility for deciding on his own to steal Ann’s car nor his responsibility for deciding to steal Ann’s car derives from the other; these are each something for which Bob is basically morally responsible.

Of these options, (ii) appears to be the least plausible. Indeed, it is difficult to imagine what reason there could be for thinking that Bob’s moral responsibility for deciding on his own to steal Ann’s car derives from his responsibility for so deciding simpliciter—particularly since it seems so natural to explain the latter in terms of the former. This is evinced by the fact that so many proponents of Frankfurt-style argument (including Mele and Robb, and even Frankfurt himself) have sought to support the judgment that agents are genuinely morally responsible for their decisions in these cases precisely by pointing out that these are decisions the agents made on their own. For this reason, (iii) too seems problematic insofar as it suggests that there is no connection between Bob’s responsibility for deciding to steal Ann’s car and his responsibility for so deciding on his own. Simply on the face of it, then, (i) appears to provide the most, if not only, plausible way of thinking about the relationship between Bob’s responsibility for deciding on his own to steal Ann’s car and his responsibility for so deciding simpliciter.

From (2) and (3) it follows that

(4) A counterexample to PRAP must be a case in which a person had no robust alternatives to deciding on his own to act as he did.
As noted earlier, here, I want to grant that having robust alternatives involves everything Fischer, McKenna, and Pereboom say it does except for the requirement that they (at least partly) consist in actions. Thus, let us assume that

(5) An alternative is robust if and only if it is (i) morally significant, (ii) deliberatively significant, (iii) understood to be responsibility-precluding, and (iv) within an agent’s voluntary control.

For the sake of simplicity, most (if not all) of the debate surrounding PAP, Frankfurt-style cases, and robustness has limited the focus to cases involving blameworthiness. Following that custom, let us assume that the relevant actions in these cases are morally bad actions for which the agents are (purportedly) not only morally responsible but blameworthy.

The next major step is to show that the alternative of voluntarily omitting to decide on one’s own to do that which one knows to be wrong satisfies the conditions for robustness identified in (5). To begin, I take it to be noncontroversial that

(6) Omitting to decide on one’s own to do that which one knows to be wrong is a morally significant alternative to so deciding on one’s own (in all Frankfurt-style cases).

Arguably, omitting to decide on one’s own to do that which one knows to be wrong (all things considered) is always a morally significant alternative to (because morally better than) deciding on one’s own to do that which one knows to be wrong. The one possible exception to this might be cases in which a person knows that her deciding to A is unavoidable, regardless of whether she decides on her own to A (say, because she knows that she is in a Frankfurt-style scenario). Of course, in such cases, it’s no longer obvious that performing the action would be morally wrong, as opposed to just morally bad. Regardless, since it is an essential feature of Frankfurt-style cases that the agents neither believe nor ought to believe that there are counterfactual interveners about, ready to force them to decide to A unless they do so on their own, these considerations do not apply in Frankfurt-style cases.

I also take it to be evident that
(7) Omitting to decide on one’s own to do that which one knows to be wrong is a deliberatively significant alternative to so deciding on one’s own (in all indeterministic Frankfurt-style cases).

Assuming that agents in these cases are morally responsible for what they do (as we are to suppose), omitting to decide on their own to do that which they know to be wrong is a morally significant alternative that it is reasonable for us to expect them to consider as an alternative to what they actually do.\(^{54}\)

Similarly, any agent satisfying the epistemic conditions for morally responsible agency will understand that she can avoid being (basically) morally responsible for doing something by simply omitting to do so on her own; for any morally competent agent will understand that one is not (basically) morally responsible for actions one is forced to perform or for actions one does not perform at all, and these are the only possibilities when one omits to perform an action on one’s own.\(^{55}\) Thus,

(8) Omitting to decide on one’s own to perform an action is understood to preclude one from being (basically) morally responsible for so deciding (in all Frankfurt-style cases).

Next, it is trivially true that

(9) Voluntarily omitting to decide on one’s own to do that which one knows to be wrong is within one’s voluntary control whenever one voluntarily omits to do so.

\(^{54}\) For an extended discussion of this point, see Chapter 3.

\(^{55}\) Certainly there may be cases in which agents can be morally responsible for the performance of actions that they are forced to perform, provided that they are morally responsible for the fact that they are forced to perform them. For instance, if Jones freely arranged for Black to force him to perform a given action, Jones could be morally responsible for the performance of this action even though this is not an action he performs on his own. In all such cases, however, agents will be only derivatively morally responsible for the performance of those actions they do not perform on their own. Since PAP is a thesis about basic moral responsibility, these cases are irrelevant to the present point.
From (5)-(9), it follows that

(10) Voluntarily omitting to decide on one’s own to do that which one knows to be wrong is a robust alternative to so deciding on one’s own (in all indeterministic Frankfurt-style cases).

That is to say, voluntarily omitting to decide on one’s own to do that which one knows to be wrong is an alternative which, *when it remains open to agents in indeterministic Frankfurt-style cases*, is a distinctly robust alternative possibility. (10) should *not* be understood as saying that this alternative *in fact* remains open to agents in all indeterministic Frankfurt-style cases. (That claim comes next.)

The final major step in the argument is to show that

(11) It is an essential feature of indeterministic Frankfurt-style cases that it is possible for agents in these cases to voluntarily omit to decide on their own to do that which they know to be wrong.

It is precisely because it is possible for these agents not to perform the relevant actions on their own that the ensuring mechanisms are there in the first place; otherwise they would be superfluous. In indeterministic cases, this option is available right up until the moment of choice, within the scope of agents’ voluntary control. The only way to prevent an agent from voluntarily omitting to decide on her own to perform a given action without causally determining what she does is to render her unconscious or otherwise not in control of her behavior, thereby precluding her responsibility. Thus, provided they are morally responsible for what they do, agents in these cases will be able to voluntarily omit to decide on their own to do so.

To be clear, I allow that it is possible to construct (indeterministic) Frankfurt-style cases in which agents are (basically) morally responsible for deciding to perform some action despite

---

\(^{56}\) Returning this alternative to within the scope of agents’ voluntary control was one of the results of the general move away from Frankfurt-style cases that relied on determinative prior signs (see McKenna and Widerker 2003, pp. 6-10).
being unable to intentionally omit to decide on their own to perform that action. As I have already argued, however, an omission need not be intentional in order to constitute a robust alternative possibility. 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. Just as one can knowingly and voluntarily perform an action and yet not perform it intentionally, one can knowingly and voluntarily omit to perform an action without intentionally omitting to perform it.

From (10) and (11), it follows that

(12) It is an essential feature of indeterministic Frankfurt-style cases that agents in these cases have robust alternatives to deciding on their own to act as they do.

And from (4) and (12), it follows that

(13) Indeterministic Frankfurt-style cases are theoretically incapable of serving as genuine counterexamples to PRAP.

So long as it remains open to agents in these cases to knowingly and voluntarily omit to decide on their own to do that which they know to be wrong, they will continue to possess robust alternative possibilities. And since there is no way to eliminate the possibility of agents knowingly and voluntarily omitting to decide on their own to act as they do without undermining the judgment that they are morally responsible for their actions, agents in these cases will continue to possess robust alternative possibilities as long as they are morally responsible for what they do. Thus, either agents are not morally responsible for what they do in these cases or

57 It is generally accepted that in order for an action to be intentional it must be suitably related to an intention of a certain sort (e.g., an intention to perform that action, or an intention to try to perform that action). Plausibly, the same is true of intentional omission (see Clarke 2010). Insofar as this is correct, (indeterministic) Frankfurt-style scenarios can eliminate agents’ ability to intentionally omit to decide on their own to act as they do simply by blocking them from coming to have the relevant intention(s).

they continue to possess robust alternative possibilities. Either way, these cases will fail to falsify PRAP and so will not be useful in mounting a successful indirect Frankfurt-style attack on PAP.

4.6 An Objection and a Reply

David Widerker (2003) has suggested that omitting to decide on one’s own to act as one does is not a robust alternative possibility in (indeterministic) Frankfurt-style scenarios because it would be morally unreasonable to expect agents in these cases to have availed themselves of this alternative, even if it was within their voluntary control to do so. To omit to make a given decision on one’s own, he says, means either not to make it at all or to be forced to make it. Thus, since it is not possible for agents in Frankfurt-style cases to avoid making the decisions they make, the expectation that they should have omitted to do so on their own “reduces to” the expectation that they should have made it the case that they are forced to so decide (p. 65).

According to Widerker, however, “A person can reasonably be expected to make it the case that \( p \) only if he knows (or should have known) or has good reasons to believe that he should make it the case that \( p' \)” (p. 65). Thus, given that agents in Frankfurt-style cases do not know (nor should they have known) that they are in situations where they will be forced to make the relevant decisions if they do not do so on their own, it would be morally unreasonable to expect them to make it the case that they are forced to make the decisions they make.\(^5^9\)

The problem with this objection is that it rests on an equivocation between two different senses of (moral) expectation. Like ‘know’, ‘expect’ is an intensional verb which allows for

\(^5^9\) Widerker couches his discussion of this alternative possibility in terms of an agent’s not making (or bringing it about that she does not make) the relevant decision freely, whereas I have been describing this as an agent’s omitting (or seeing to it that she does not make) the relevant decision on her own, but this difference in terminology does not appear to amount to a substantive difference. We are both addressing the same alternative.

\(^6^0\) The kind of expectation at issue here is moral in nature and belongs to a class of what we might call normative expectation. To subject someone to an expectation of this sort is to hold her to a certain standard of behavior. It should not be confused with what we might call predictive or anticipatory expectation, which is a matter of believing that things will go a certain way in the future (see Mellema 1998). These two kinds of expectation are quite independent of one another and often come apart. A set of parents, for instance, might (morally) expect of their teenage son that he call them to let them know if he plans on staying out late that night even though, given his track record, they do not (predictively) expect that he will actually do this. Furthermore, it could be morally reasonable for the parents to expect this of their son even if it would not be rational for them to believe (or expect) that he will do so.
ambiguity with respect to its complements. Widerker’s argument fails to acknowledge this fact and so capitalizes on the resultant ambiguity. Once this ambiguity is resolved, the objection fades.

Consider, first, Widerker’s claim that, since it is not possible for them to avoid making the decisions they make, the expectation that agents in Frankfurt-style cases should have omitted to decide on their own to act as they do reduces to the expectation that they should have made it the case they are forced to so decide. How should we understand this? In what sense does (subjecting these agents to) the former expectation reduce to (subjecting them to) the latter? There are at least two ways to interpret this claim:

(1a) To expect of agents in Frankfurt-style cases that they omit to decide on their own to A entails subjecting them to an expectation with the content “Make it the case that you are forced to decide to A.”

(1b) To expect of agents in Frankfurt-style cases that they omit to decide on their own to A is in effect to subject them to the expectation that they make it the case that they are forced to decide to A.

(1a) seems obviously to be false. Whereas an expectation with the content “Omit to decide on your own to A” would be met in any possible world where the agents omit to make the relevant decisions on their own—regardless of whether they make it the case that they are forced to make those decisions—an expectation with the content “Make it the case that you are forced to decide to A” would be met only in cases where the agents are forced to make the relevant decisions. Subjecting agents to the former expectation need not involve, nor does it entail, subjecting agents to the latter expectation—even in cases where, as a matter of contingent fact, acting in such a way that meets the former expectation would also be sufficient for meeting the latter.

Presumably, then, (1b) is closer to what Widerker has in mind. In addition to being the only plausible alternative of the two, it is the only interpretation of this claim that can reasonably be thought to follow from the considerations Widerker offers in its favor. To better see this connection, it will be helpful to spell out precisely what it is to in effect expect of a person that
she make something the case (as opposed to in fact expecting of a person that she make something the case). Here is a proposal: To in effect expect of a person that she make it the case that \( p \) is to expect of her that she do (or omit to do) \( q \) and to subject her to that expectation under conditions in which her doing (or omitting to do) \( q \) will make it the case that \( p \). With this understanding in place, we can further clarify (1b) as follows:

\[
\text{(1b*) To expect of agents in Frankfurt-style cases that they omit to decide on their own to A is to expect of them that they omit to decide on their own to A and to subject them to this expectation under conditions in which omitting to decide on their own to A will make it the case that they are forced to so decide.}
\]

Understood in this way, (1b) appears to be uncontroversial.

Now consider Widerker’s claim that “[a] person can reasonably be expected to make it the case that \( p \) only if he knows (or should have known) or has good reasons to believe that he should make it the case that \( p \)” (p. 65). As with the first statement, there are at least two ways this can be interpreted:

\[
\text{(2a) It is morally reasonable to subject a person to an expectation with the content “Make it the case that \( p \)” only if she knows (or should have known) or has good reasons to believe that she should make it the case that \( p \).}
\]

\[
\text{(2b*) It is morally reasonable to expect of a person that she do (or omit to do) \( q \) and to subject her to that expectation under conditions in which her doing (or omitting to do) \( q \) will make it the case that \( p \) only if she knows (or should have known) or has good reason to believe that she should make it the case that \( p \).}
\]

Here, the precise nature of the problem with Widerker’s argument begins to come into focus. Nothing follows from the conjunction of (1b*) and (2a) since these statements are about different things. The same goes for (1a) and (2b*). Thus, even assuming that (2a) is correct, this will be helpful in showing that omitting to decide on one’s own to do that which one knows to be
wrong is not a robust alternative in Frankfurt-style cases only when taken together with (1a). As noted earlier, however, (1a) is pretty plainly false.

The only other valid inference to the judgment that it is morally unreasonable to expect of agents in Frankfurt-style cases that they omit to decide on their own to act as they do is from (1b*) and (2b*). Like (1a), however, (2b*) should be rejected. Consider, for instance, a case in which Lois has promised Clark that she will meet him for lunch on Monday. All else being equal, it is morally reasonable to expect of Lois that she keep her promise to meet Clark for lunch. Now suppose that, looking for ways to amuse himself, Jimmy has wagered Clark twenty dollars that Lois will fail to show up. As a result, if Lois keeps her promise to meet Clark for lunch, then, unbeknownst to her, she will thereby make it the case that Clark wins twenty dollars. According to (2b*), this entails that it is morally unreasonable to expect of Lois that she keep her promise to meet Clark for lunch. But surely this is incorrect. Given that Lois is unaware (and had no good reason to be aware) of Clark’s wager with Jimmy, it might well be morally unreasonable to expect of her that she make it the case that Clark wins twenty dollars (that is, to subject to her to an expectation with the content “Make it the case that Clark wins twenty dollars”). Furthermore, since, as a matter of contingent fact, keeping her promise to meet Clark for lunch will thereby make it the case that Clark wins twenty dollars, it is plausible that to expect of Lois that she keep her promise to Clark is in effect to expect of her that she make it the case that Clark wins twenty dollars, in the sense identified in (1b*) (namely, that this is to expect of Lois that she keep her promise to meet Clark for lunch and to subject her to that expectation under conditions in which her meeting Clark for lunch will make it the case that Clark wins twenty dollars). Nevertheless, it is still morally reasonable to expect of Lois that she keep her promise to meet Clark for lunch. After all, it is surely reasonable to expect this of her in the original version of the case without the wager, and it is utterly mysterious how the mere addition of Jimmy’s wager with Clark could do anything to change this.\(^{61}\)

\(^{61}\) Similarly, we could suppose that, by keeping her promise to meet Clark for lunch, Lois would make it the case that she is the seventeenth person to sit in a particular chair that day, or that the temperature in the restaurant is 0.2 degrees higher than it otherwise would have been, or that any number of other equally obscure states of affairs obtain. It seems passing strange to think that the reasonableness of expecting of Lois that she keep her promise to Clark would hinge on whether she is aware (or ought to be aware) of any of these things.
The same is true of agents in indeterministic Frankfurt-style cases. Assuming they are morally responsible for the decisions they make in these cases, it is morally reasonable to expect of them that they omit to decide on their own to do that which they know to be wrong—just as it would be reasonable to expect this of them in ordinary (non-Frankfurt-style) scenarios—which it is within their (direct) voluntary control to do. Of course, since, as Widerker points out, it is a crucial feature of these cases that the agents do not know (nor should they have known) that they will be forced to make these decisions unless they do so on their own, it might well be that it is unreasonable to expect of them that they make it the case that they are forced to so decide. But this does nothing to diminish the reasonableness of expecting them to omit to decide on their own to do that which they know to be wrong.

4.7 Conclusion

Toward the end of the 1968 film, *The Lion in Winter*, there is a scene where Richard, Geoffrey, and John have been locked in the dungeon by their father, King Henry II, and are waiting to be executed. Eventually, upon hearing footsteps coming down the stairs, and believing they had finally seen sunlight for the last time, Richard stands to his feet, defiantly declaring, “He’ll get no satisfaction out of me. He isn’t going to see me beg.” “Why, you chivalric fool,” Geoffrey chides, “as if the way one fell down mattered.” “When the fall is all there is,” Richard replies, “it matters.”

Critics of the flicker strategy have argued that the alternatives that remain open to agents in indeterministic Frankfurt-style examples, including the ability to omit to perform the relevant actions on their own, are unable to do the work set out for them, that they are not the kinds of alternative on which proponents of the ability to do otherwise have traditionally been focused, and that they are too diminutive to matter. Granted, generally believing that it remains up to us not only how we perform the actions we perform but whether we perform them at all, this is not the full-fledged freedom with which we are normally concerned and which we commonly take ourselves to possess. Still, when this flicker is all there is—when the ability to knowingly and voluntarily omit to do that for which one is basically morally responsible is all the freedom that remains—it matters.
CHAPTER FIVE

ALTERNATIVE CONCEPTIONS OF MORAL RESPONSIBILITY

5.1 Introduction
Whether being morally responsible for what one has done requires that one could have done otherwise largely depends on just what it is to be morally responsible for something. In the debate surrounding Frankfurt-style cases, a great deal of time and effort has been spent clarifying what precisely is involved in an agent’s having had the kind of ability to do otherwise referenced in PAP. This includes answering questions concerning the exact time at which an agent must have been able to do otherwise, whether doing otherwise must be intentional, whether doing otherwise must involve an action, whether it must be a completed action or whether merely initiating an action would suffice, along with a host of other related issues. Relatively little attention has been paid, however, to the nature of moral responsibility itself. Indeed, in Frankfurt’s original article (in what may be my favorite footnote in a philosophy paper), he writes:

The two main concepts employed in the principle of alternate possibilities are “morally responsible” and “could have done otherwise.” To discuss the principle without analyzing either of these concepts may well seem like an attempt at piracy. The reader should take notice that my Jolly Roger is now unfurled. (1969, p. 835)

Such an analysis has been unnecessary for much of this debate since parties on both sides seem mostly to have been operating with a similar notion of moral responsibility—namely, a desert-entailing conception of responsibility, according to which one who is morally responsible is thereby capable of deserving blame or praise and punishment or reward for things one has done. In recent years, however, there have been numerous attempts to reconceive the nature of moral responsibility in a way that straightforwardly implies that whether an agent is morally
responsible for something he has done has nothing to do with whether he had alternative possibilities.

My aim in this chapter is both to draw attention to and to raise some worries about a few of these conceptions of moral responsibility. In what follows I consider two general ways that some have attempted to revise the traditional understanding of what it is to be morally responsible, each of which implies that the ability to do otherwise is a non-issue. The first (and more common) way is to defend a deflationary account of moral responsibility, according to which responsibility does not entail that one is (or even can be) deserving of any amount of blame or praise, punishment or reward for things one has done. The second way is to propose an inflationary account of moral responsibility, according to which responsibility entails not only that one is deserving of some amount of praise or blame, punishment or reward but that one is (or at least could be) deserving of an infinite amount of punishment or reward. One problem that is shared by each of these conceptions of moral responsibility, I argue, is not merely that they imply alternative possibilities are not required for moral responsibility but that it is unclear why anyone would think that bearing these kinds of responsibility has anything to do with the ability to do otherwise. Insofar as a given conception is unable either to give an adequate account of the force and fairness typically associated with judgments of moral responsibility or to allow for the intelligibility of the longstanding debate over the necessity of alternative possibilities, so much the worse for that conception of responsibility.

5.2 Deflationary Accounts of Moral Responsibility

Generally speaking, an account of any kind will be either deflationary or inflationary only in comparison to some other account. By characterizing the following accounts of moral responsibility as deflationary, I am contrasting them with what I take to be the traditional account (or, perhaps, accounts) of moral responsibility, one which is capable of grounding ascriptions of desert. Now, while I am also inclined to say that any account of moral responsibility that is unable to ground ascriptions of desert is not only deflationary but overly deflationary, whether and when such an account is overly deflationary is somewhat tricky business. Presumably, an account of moral responsibility will be overly deflationary when it fails to adequately account for the force and fairness typically associated with judgments of moral responsibility. As a result, its scope will be too broad; it will count as being morally responsible agents who, according to our
best considered judgments, ought not so to be counted. In this sense, each of the following accounts, I shall argue, looks to be not only deflationary but overly deflationary.

It is important to point out, however, that this way of thinking about things presupposes that these accounts are correctly understood as descriptive accounts of moral responsibility, whose aim is to adequately characterize a given property (or set of properties) in the real world. Moreover, it also presupposes that the property (or set of properties) that these deflationary accounts intend to describe is the very same property (or set of properties) that the traditional account of moral responsibility intends to describe. It may be, however, and there is some evidence to suggest, that these accounts are meant to be more definitional in nature. If so, then, rather than being competitors to the traditional account of moral responsibility, these accounts will be properly understood as describing different kinds or senses of moral responsibility. Consequently, ascriptions of this kind of moral responsibility may well be compatible with considered judgments of non-responsibility in the traditional, desert-entailing sense, thereby eliminating the reason to classify these deflationary accounts as overly deflationary in the sense described above. What is most important for present purposes is that, assuming that PAP is meant as a thesis about desert-entailing moral responsibility, insofar as these deflationary accounts of moral responsibility are intended as definitional accounts describing different kinds or senses of moral responsibility, the irrelevance of alternative possibilities to the kinds of moral responsibility characterized by these accounts will not entail the falsity of PAP. This last point is often obfuscated, in my view, by the use of identical terminology (such as ‘responsible’, ‘guilt’, and ‘blame’) together with a failure to indicate with sufficient clarity when and how this terminology is being used in different ways. My chief aim in this chapter is to contribute to the elimination of this obfuscation.

5.2.1 Smart on Responsibility

According to J.J.C. Smart (1961), in general, to praise or dispraise a thing is just to grade it, that is, to evaluate it according to some standard. A thing is praised when it receives a positive evaluation and dispraised when it receives a negative evaluation, just as a student might be praised for her mathematical skills and an athlete might be dispraised for his incompetence on the field. And the same, he says, is true of moral qualities and actions:
Just as we may praise or dispraise a woman for her figure, a footballer for his fleetness or slowness of foot, a lecturer in philosophy for his intelligence or lack of intelligence, and a writer for clarity or obscurity, so naturally enough, we may praise or dispraise a man for his honesty or dishonesty, truthfulness or untruthfulness, kindness or unkindness, and so on. In this sense of ‘praise’ we may praise moral qualities and moral actions in exactly the same way as we may praise beauty, intelligence, agility or strength. (1961, p. 304)

Of course, although we may dispraise a person for his ugliness, we do not typically blame him for it. The reason for this, Smart maintains, is that, in addition to grading, to blame a person for something implies that it is both open to modification and something for which she is responsible in the sense that it is the result of certain choices she made, though not necessarily in any deeply metaphysical sense. Aside from this, however, blame should be viewed in just the same way as dispraise. “Blame in this sense,” he says, “can be just as dispassionate as dispraise of a woman’s nose: it is just a grading plus an ascription of responsibility” (p. 305).

On Smart’s view, then, to be praiseworthy or blameworthy (and so morally responsible) is just for us to be causally responsible for something that is both modifiable and open to being graded according to certain moral standards. Obviously, if Smart were correct in thinking that this is what it is to be morally responsible for something one has done, there would be no reason to hold that the ability to do otherwise is a necessary condition for moral responsibility. Although Smart’s conception of moral responsibility is widely rejected as being overly deflationary, recently others have defended accounts of moral responsibility that are not all that dissimilar to Smart’s, either in their deflationary nature or in their implications concerning the (ir)relevance of alternative possibilities.

5.2.2 Scanlon on Responsibility

In Chapter 6 of his book, What We Owe to Each Other, T. M. Scanlon (1998) sets out to offer an account of moral responsibility that is compatible with his overall contractualist moral theory. Scanlon distinguishes between two senses of moral responsibility: responsibility as attributability and substantive responsibility. Whereas responsibility as attributability concerns whether a person is properly subject to moral appraisal, such as praise or blame, substantive responsibility has to do with whether a person is properly subject to various burdens or
obligations. It is the former sense of responsibility on which I wish to focus. Here I am particularly interested in Scanlon’s claim that psychopaths—those with an inability to recognize moral reasons and for whom moral reasons cannot have any motivational force—can be properly subject to moral appraisal. This is significant not only because of the difficulties it raises for Scanlon’s view but because it serves to highlight the deflationary nature of the conception of moral responsibility he defends.

According to Scanlon, morality is fundamentally a system of co-deliberation in which moral reasoners attempt to arrive at principles that each of them could be asked to use to govern their deliberation and accept as a basis of criticism. For an action that one performs to be wrong, on this view, is just for it to be disallowed by those principles that no one (properly motivated) could reasonably reject (1998, pp. 268-269). Furthermore, acting wrongly does not require that one knowingly or intentionally act in a way that is not allowed by such principles, though this would certainly be sufficient to qualify what one does as wrong. One also acts wrongly when one simply fails to notice that what one does violates such principles, provided that this failure is the result of faulty self-governance. The kind of self-governance at issue here is has less to do with the regulation of behavior by conscious (and perhaps intentional) acts of will, such as decisions, and more to do with the regulation of one’s judgments by one’s perception of reasons. As Scanlon puts it, “A person governs herself in the sense required if she is sensitive to the force of reasons and to the distinctions and relations between them and if her response to these reasons generally determines her subsequent attitudes and actions” (p. 281).

To subject someone to moral criticism, then, is partly to claim that one has governed oneself in a way that is not allowed by those principles that no one could reasonably reject (p. 268). Beyond this, to blame or criticize someone is to demand that one either justify oneself, by showing that the claim is mistaken, or acknowledge that one’s self-governance has been faulty and must be modified. Such a demand makes sense, Scanlon says, only if that in virtue of which one is being criticized is something for which one is broadly responsible, which is to say, under the control of one’s judgment-sensitive attitudes. One might feel embarrassed about one’s excessively large ears or ashamed of one’s intellectual abilities, for instance, but since these are not things for which one is responsible in the broad sense, “there is nothing for one to ‘take back’ or modify” (p. 270). It is for this reason, he maintains, that the scope of moral criticism is limited to the judgment-sensitive attitudes of rational creatures, which, unlike physical characteristics,
are under the control of one’s reason. It is only when one either fails to notice or knowingly flouts certain considerations that, according to those principles that no one could reasonably reject, count against one’s actions that one is blameworthy for what one does. But whenever one’s self-governance is faulty in one of these two ways, one is the appropriate subject of moral criticism (pp. 269-272). Thus, it is both necessary and sufficient for being morally responsible for what one does, on Scanlon’s view, that one who possesses the general capacity to respond to and assess reasons governs oneself in a way that allows one’s actions to be attributable to oneself.

Since, as we have seen, the kind of self-governance Scanlon has in mind is a matter of whether one’s judgments are sufficiently sensitive to reasons in the right sort of way and has nothing to do with the exercise of an agent’s will, governing oneself in this way does not require the possession of alternative possibilities. Nor does having the general capacity to respond to and assess reasons. Insofar, then, as one accepts the account of moral responsibility Scanlon proposes, one should reject PAP. But is Scanlon’s account one that we should accept?

Scanlon states that one of his aims in this chapter is to argue that his conception of moral appraisal “offers plausible accounts both of the range of the things that can be objects of moral assessment and the force of the content of our notions of moral blame and guilt” (p. 267). It is perhaps somewhat surprising, then, that he is so willing to accept the consequence that psychopaths turn out to be morally responsible for their actions according to his account (p. 284). Psychopaths are those who, though subject to judgment-sensitive attitudes and capable of practical deliberation in general, are unable to see the force of, and so are unable to be motivated by, morally relevant reasons. Such a person might see that a particular course of action would cause harm to others but is incapable of seeing that this constitutes a reason for him to refrain from that course of action. According to many (if not most) philosophers, the fact that psychopaths are incapable of recognizing and responding to moral reasons is sufficient to preclude them from being morally responsible for their actions. Insofar as one thinks this is correct, one may resist Scanlon’s conception of moral appraisal on the grounds that the range of

---

62 See, for example, Fischer and Ravizza 1998 (pp. 76-81), Wallace 2002 (pp. 441-445), and Dworkin 2002 (pp. 478-482).
things that can be the appropriate objects of such appraisal is implausibly broad—at least, to the extent that Scanlon’s view does in fact require seeing psychopaths as subject to moral appraisal.

On the face of it, there are at least two reasons for thinking that psychopaths are not properly subject to moral appraisal, one of which is provided by Scanlon’s own view. Recall that, according to Scanlon’s conception of moral appraisal, part of what it is to blame a person for what she has done is to call on her to either justify her judgment-sensitive attitudes or acknowledge guilt and modify them accordingly. Indeed, it was for this reason that Scanlon argued that moral appraisal applied only to the judgment-sensitive attitudes of rational creatures in the first place. The problem, of course, is that psychopaths are incapable of either justifying their attitudes or acknowledging their guilt and modifying them since to do so would require that they recognize the moral reasons in terms of which, and in response to which, this justification and modification are to occur. It therefore seems unreasonable to demand of them that they do so.63

Scanlon denies that the fact that psychopaths are incapable of recognizing and responding to the special force of moral reasons precludes them from being blameworthy for their failure to properly incorporate those reasons into their practical deliberation. “If a person holds a certain judgment-sensitive attitude,” he says, “then, because this attitude is in principle sensitive to and dependent on his judgment, it is appropriate in a general sense to ask him to defend it or disown it” (1998, p. 289). Here his view seems to be that, even if a person is incapable of modifying his judgment-sensitive attitudes in response to moral reasons, it is appropriate to demand of him that he do so because these attitudes are in principle the sort of thing that is sensitive to reasons—at least, in a general sense. But what does it mean for it to be appropriate “in a general sense”? The question, presumably, is whether it is appropriate to demand that a particular person either justify or modify a particular attitude in response to particular moral reasons when that person has a particular inability to do so. And Scanlon’s answer, surprisingly, appears to be yes. Perhaps even more surprising than his answer, however, is the reason that he gives for it. It is not at all obvious in what way the demand for a person who is unable to recognize and respond to reasons of a certain sort—namely, moral reasons—is made appropriate in virtue of the fact that these attitudes belong to a class of states that are in principle, though not in him, sensitive to judgment.

63 Wallace (2002) has also raised this objection.
Another thing that is not immediately clear is whether in fact Scanlon is, in virtue of his conception of moral appraisal, theoretically committed to the view that the psychopath is morally responsible for what he does—thus making his insistence on this point all the more curious. In addition to the claim that the psychopath’s attitudes are in principle sensitive to judgments, Scanlon offers a further consideration in support of the view that psychopaths are properly subject to moral criticism which may provide a clue as to why Scanlon takes this position. According to Scanlon, even if a person is

unable to see the force of some reason that counts against [an attitude for which he is blamed], this does not alter the fact that the attitude and the judgment that it is warranted are properly attributable to him. Any errors involved in these attitudes are also attributable to him, and he is therefore properly criticized for holding them. (This is certainly the view we take with respect to the judgment-sensitive attitudes of other kinds, such as beliefs about mathematics or strategic judgment in chess. (1998, p. 289)

Here there are two things that Scanlon says that are particularly worth noting. The first thing to note is the background assumption that it is sufficient for being properly subject to moral criticism for what a person does that the attitudes expressed in his actions are both attributable to him and faulty, a point he makes explicit earlier (p. 287). The second thing to note is the suggestion that, where moral appraisal is concerned, we ought to view judgment-sensitive attitudes about moral concerns just as we view similar attitudes concerning mathematical issues.

At this point we may begin to suspect that the conception of moral criticism just presented is strikingly similar to Smart’s conception of blame as grading, and so fails to account for the special force that blame is typically taken to have. Scanlon insists that his account of moral responsibility differs from Smart’s view (which he considers to be overly deflationary) in two important respects. First, moral criticism on his view applies only to a person’s judgment-sensitive attitudes, regarding which it calls on him to either defend or modify and withdraw. Second, moral criticism differs from other unwelcome negative evaluations in that “the failings to which it calls attention have a particular kind of significance for the agent’s relations with other people” (1998, pp. 275-276).
Do these two features show that Scanlon’s view is significantly different from Smart’s view? The first distinction merely notes a difference in the scope of moral criticism, by restricting it to those attitudes that are in general sensitive to reasons. Thus, it is the second distinction which is left to account for the special force of moral appraisal (over and above the sort contained in Smart’s account). It may well be true that negative evaluations of our judgment-sensitive attitudes, such as our being racist or ungrateful toward others, carry with them implications for our relations with others in a way that judgments about our intelligence and physical appearance do not. But just how much force this actually affords ascriptions of praise and blame, above and beyond that which is present in Smart’s view of moral grading, is unclear. It certainly does not seem that this distinction amounts to nearly as great a difference as Scanlon suggests, indicating that his view is in fact not all that dissimilar to Smart’s extremely deflationary account of moral appraisal. The more deflationary his account of moral appraisal, the more plausible it is to say that psychopaths can be morally responsible for their actions. At the same time, however, the more deflationary his account of moral appraisal becomes, the less plausible it is on the whole, since it seems less and less able to account for the special force that blame is typically taken to have.

5.2.3 Smith on Responsibility

In a series of recent papers, Angela Smith (2005, 2007, 2008) has developed and defended a version of Scanlon’s account of moral responsibility, which she refers to as the rational relations view. “To say that an agent is morally responsible for some thing, on this view,” she says, “is to say that that thing reflects her rational judgment in a way that makes it appropriate, in principle, to ask her to defend or justify it” (2008, p. 369). Like Scanlon, Smith sees moral criticism (blame) as involving a demand that the one being blamed either defend or modify that for which one is blamed. Indeed, it is this feature of blame, on her view, that “primarily distinguishes moral criticism from other forms of negative description or assessment” (p. 381). As a result, Smith also follows Scanlon in holding that the primary objects of moral responsibility are agents’ judgment-sensitive attitudes since these are the sort of thing it can, in principle, make sense to be asked to justify or defend. Although we can certainly be morally responsible for actions we perform (and fail to perform) as well as the consequences of those
actions (and omissions), in her view it is only because they reflect our judgment-sensitive attitudes that they carry the moral and interpersonal significance that they do.

Smith contends that her “non-volitionalist” account of moral responsibility is preferable to what she labels “volitionalist” accounts. According to volitionalist accounts, agents are morally responsible only for things that fall within the scope of their voluntary control. Volitionalist accounts come in different flavors. Whereas some volitionalists restrict moral responsibility to that which can be traced to prior decisions an agent has made, others require only that an attitude be (voluntarily) endorsed by the agent once it appears. Still other volitionalists allow that agents can be morally responsible for anything that is susceptible to their voluntary control, even if it does not have its origin in any voluntary act and is not explicitly endorsed (2008, pp. 381-382; see also 2005, pp. 238-240). Now, the reason that defenders of PAP maintain that a person can be morally responsible for what he has done only if he could have done otherwise has to do with concerns about control. If, as Smith argues, one can be morally responsible for things that lie entirely beyond one’s voluntary control, then responsibility will certainly not require dual voluntary control—that is, the ability either to voluntarily pursue some course of action or to voluntarily refrain from doing so. Thus, like Scanlon’s account of moral responsibility, Smith’s rational relations view implies that PAP should be rejected.

The main argument that Smith offers in favor of accepting her non-volitionalist, rational relations view of moral responsibility is that it “makes better sense of our ordinary intuitions about the conditions under which a person is responsible for her attitudes” (2005, p. 240). Here, Smith is particularly interested in spontaneous mental attitudes over which one has not exercised any control. First, we regularly use a person’s unreflective patterns of awareness—what a person notices and neglects to notice—as the basis for moral appraisal of that person. She offers the following example:

I forgot a close friend’s birthday last year. A few days after the fact, I realized that this important date had come and gone without my so much as sending a card or giving her a call. I was mortified. What kind of friend could forget such a thing? Within minutes I was on the phone to her acknowledging the fault and offering my apologies.
But what, exactly, was the nature of my fault in this case? After all, I did not consciously *choose* to forget this special day or deliberately *decide* to ignore it. I did not *intend* to hurt my friend’s feelings or even *foresee* that my conduct would have this effect. I just forgot. It didn’t occur to me. I failed to notice. And yet, despite the apparent involuntariness of this failure, there was no doubt in either of our minds that I was, indeed, responsible for it. Although my friend was quick to pardon my thoughtlessness and to dismiss it as trivial and unimportant, the act of pardoning itself is simply a way of renouncing certain critical responses which it is acknowledged would, in principle, be justified. (2005, p. 236)

According to Smith, the fact that she did not decide to forget her friend’s birthday is not enough to preclude this from serving as a legitimate basis for moral criticism—which, recall, is just what it is to be morally responsible on her view. The reason that she would be legitimately subject to criticism for her failure to remember her friend’s birthday, in her view, is that “we take there to be some sort of rational connection between what we notice and what we evaluate or judge to be important or significant” (p. 244). When she forgets her friend’s birthday, this indicates that she does not judge her friend or her friend’s feelings to be as important as she should.

Smith appeals to other examples to show that, in addition to what we notice and neglect to notice, we also judge people to be morally responsible for what occurs to us as well as for involuntary reactions. If, for instance, it occurs to an individual that he could solve a serious problem with a business deal by having someone killed, this too is something for which he can be blameworthy—even if he immediately rejects the idea as morally impermissible:

The reason we continue to feel uneasy with such a person, I submit, is that we assume that for such an appalling alternative even to “come into thought” as a serious option, this person must either not accept or not be fully committed to, the evaluative judgment that killing others is not acceptable way of furthering one’s own interests. At the very least, we may feel, he must place shockingly little value on human life, if he is willing to consider sacrificing it, even momentarily, for business or political ends. (p. 247)
It is because of the way that what occurs to us is rationally connected to our judgments that these are things that we can be called on to justify or defend. The same is true, she says, of spontaneous desires (say, to punch another person in the face) and involuntary reactions (such as jealousy or amusement) we might have (pp. 247-250). Even though they are neither chosen nor within our direct voluntary control, many involuntary responses “provide an important indication of a person’s underlying moral commitments of who he is, morally speaking” (p. 242).

Smith argues that these examples show that volitional accounts of moral responsibility “fail utterly to account for our actual practices of self- and other-assessment, which do not seem to be sensitive to these issues about choice or voluntariness” (2005, p. 257). But this conclusion is not supported by the argument she provides. In evaluating the many examples she puts forward of cases in which people are thought to be morally responsible for their involuntary reactions and desires, for what occurs to them, and for what they notice and neglect to notice, Smith is quick to stress that these things are neither the result of prior decisions nor are they things that lie within our direct voluntary control. Even supposing she is right about all this, this does not show that we ought to reject volitionism. As Smith herself points out, there are different kinds of volitionalism. Whereas some volitional accounts state that the objects of moral responsibility must be decisions or within the subject’s direct voluntary control, others require only that they are within the subject’s indirect voluntary control (or, as Smith puts it, susceptible to voluntary control). Furthermore, all of the examples she offers involve things that are within the subjects’ indirect voluntary control. Thus, while Smith may well be successful in showing that one can be morally responsible for things that lie beyond the scope of one’s direct voluntary control, the examples she offers fail to provide any reason to reject a volitionalist account of moral responsibility because she fails to provide any reason for thinking that agents can be morally responsible for things that are not even indirectly within their voluntary control.

5.3 Inflationary Accounts of Moral Responsibility

One way to make practice and athletic ability irrelevant to success is simply to lower the bar to the point where no practice or athletic ability is required. Another way to make these things irrelevant is to raise the bar so high that no amount of training or physical prowess could possibly result in one’s clearing it. We have seen how deflationary accounts of moral responsibility have been used to attempt to settle (in the negative) the question of whether the
ability to do otherwise is of central importance to moral responsibility. Inflationary accounts of moral responsibility can also be used to this effect.

5.3.1 Heaven-and-Hell Responsibility

The clearest illustration of how an inflationary account of moral responsibility can be used to show that alternative possibilities are irrelevant can be found in Galen Strawson’s (1986, chap. 2; 1989, 1994, 2000, 2002) argument for the claim that moral responsibility is impossible. Strawson has argued that the common conception of moral responsibility—which he calls “heaven-and-hell responsibility”—is such that one can bear responsibility of this sort only if one is causa sui (“self-caused”). Since nothing can be the cause of itself, it is impossible for anyone to be morally responsible in the way widely supposed.

Strawson offers several versions of his argument to the effect that, regardless of whether determinism is true, it is impossible for us to be morally responsible for what we do. The most thorough of these is formulated as follows:

1. Interested in free action, we are particularly interested in actions that are performed for a reason (as opposed to ‘reflex’ actions or mindlessly habitual actions).

2. When one acts for a reason, what one does is a function of how one is, mentally speaking…

3. So if one is to be truly morally responsible for how one acts, one must be truly responsible for how one is, mentally speaking—at least in certain respects.

4. But to be truly responsible for how one is, mentally speaking, in certain respects, one must have brought it about that one is the way one is, mentally speaking, in certain respects. And it is not merely that one must have caused oneself to be the way one is, mentally speaking. One must have consciously and explicitly chosen to be the way one is, mentally speaking, in certain respects, and one must have succeeded in bringing it about that one is that way.
(5) But one cannot really be said to choose, in a conscious, reasoned, fashion, to be the way one is mentally speaking, in any respect at all, unless one already exists, mentally speaking, already equipped with some principles of choice, ‘P1’—preferences, values, pro-attitudes, ideals—in light of which one chooses how to be.

(6) But then to be truly responsible, on account of having chosen to be the way one is, mentally speaking, in certain respects, one must be truly responsible for one’s having the principles of choice P1 in light of which one chose how to be.

(7) But for this to be so one must have chosen P1, in a reasoned, conscious, intentional fashion.

(8) But for this, i.e., (7), to be so one must already have had some principles of choice P2, in the light of which one chose P1.

(9) And so on. Here we are setting out on a regress that we cannot stop. True self-determination is impossible because it requires the actual completion of an infinite series of choices of principles of choice.

(10) So true moral responsibility is impossible, because it requires true self-determination, as noted in (3) (1994, pp. 6-7).

If being morally responsible for what we do requires our having made an infinite number of choices, as Strawson contends, then it looks as though we will have to concede that none of us is morally responsible for anything we do. Whether this is in fact a necessary condition for moral responsibility, however, depends largely on what it is to be morally responsible. Strawson characterizes the sort of moral responsibility he has in mind in the following way:

As I understand it, true moral responsibility is responsibility of such a kind that, if we have it, then it makes sense, at least, to suppose that it could be just to punish some of us
with (eternal) torment in hell and reward others with (eternal) bliss in heaven. The stress on the words ‘makes sense’ is important, for one certainly does not have to believe in any version of the story of heaven and hell in order to understand the notion of true moral responsibility that it is being used to illustrate. Nor does one have to believe in any version of the story of heaven and hell in order to believe in the existence of true moral responsibility. The story of heaven and hell is useful simply because it illustrates, in a peculiarly vivid way, the kind of absolute or ultimate accountability or responsibility that many supposed themselves to have, and that many do still suppose themselves to have. It very clearly expresses its scope and force. (1994, p. 21)

Elsewhere, in order to emphasize the purported independence of the conception of true moral responsibility from considerations of heaven and hell, Strawson describes it as

> responsibility and desert of such a kind that it can exist only if punishment and reward can be fair or just without having any pragmatic justification, or indeed any justification that appeals to the notion of distributive justice. (2002, p. 452)

Though the latter characterization makes no mention of heaven and hell, nor of the eternal rewards and punishments with which they are closely associated, Strawson takes these characterizations to describe equivalent conceptions of responsibility. Furthermore, as Strawson sees it, the sort of moral responsibility he describes, and which he takes to have been proven impossible by the argument offered above, is one that has not only “been central to the Western religious, moral, and cultural tradition” but is “a natural part of the human moral-conceptual repertoire” (1994, pp. 8-9).

Insofar as this is correct, the implications of Strawson’s conclusion would appear to be rather significant. At the very least, it would seem to call for a serious reconsideration of many of our social and moral practices as well as our conception of ourselves as moral agents and, perhaps, even as persons.\(^{64}\) Another effect of Strawson’s argument, if sound, would be to show

\(^{64}\) Whether and to what extent our not being morally responsible agents carries implications for our conception of ourselves as persons will of course depend on what it is to be a person. For a
that the question of whether we are ever morally responsible for any of the things we do in no way hinges on whether we could have done otherwise. Regardless of whether we were sometimes able not to perform certain actions we performed, we could not have been *causa sui* and so cannot bear any moral responsibility for the things we have done; alternative possibilities are irrelevant.

### 5.3.2 Clarke’s Objection to Strawson

Randolph Clarke (2005) observes that, while Strawson’s argument itself has been the subject of much controversy, the particular (and, in his view, “quite striking”) conception of moral responsibility it employs has drawn very little consideration. Closer examination of this notion, he argues, reveals that, even if the argument succeeds, the conclusion is not quite as significant as it might at first appear. 65 Although I shall argue that Clarke’s particular line of objection is ultimately unsuccessful, it is, I think, of the right spirit: Even if Strawson is correct to maintain that heaven-and-hell responsibility as he describes it is in fact impossible, there is an important kind of responsibility with which we are deeply concerned that is not obviously threatened by Strawson’s argument. Furthermore, determining whether we ever possess alternative possibilities of the relevant sort might well be central to our ability to bear this kind of moral responsibility even if it is irrelevant to the particular kind of responsibility Strawson describes.

---

view that sees morally responsible agency as essential to personhood, see Fischer and Ravizza 1998 (esp. pp. 15-16).

65 Elsewhere in the same paper, Clarke also argues that Strawson’s argument in fact fails to adequately support even its stated conclusion on the grounds that it is missing at least one crucial premise, namely:

(P) When you do what you do because of the way you are, it is not possible for it to be up to you whether if you are that way, in certain crucial mental respects, then you perform that action (2005, p. 19).

Without P it does not follow from (2) the claim that, when acting for a reason, what one does is a function of how one is in certain crucial mental respects that (3) one is truly morally responsible for how one acts only if one is truly responsible for how one is in certain crucial mental respects. P, however, is highly contentious. And although it is implicitly assumed in his argument, Strawson offers no defense of P to those who would reject it. Without such a defense, Clarke maintains, Strawson’s argument seems to carry little force.
Clarke argues that Strawson treats as equivalent two distinct conceptions of moral responsibility. These conceptions can be characterized as follows:

R1 True (or ultimate) responsibility is of such a kind that it can exist if and only if punishment and reward can be fair or just without having any pragmatic justification.

R2 True (or ultimate) responsibility is of such a kind that it can exist if and only if infinite punishment and reward can be fair or just without having any pragmatic justification.

Clarke insists that, although Strawson takes the conception of moral responsibility characterized by R1 to be equivalent to that characterized by R2, they “manifestly [are] not the same thing, for deserving punishment or reward does not entail deserving (nor even that one could deserve) infinite punishment or reward” (2005, p. 22). Thus, even if one were willing to grant that there is no sort of freedom possible for human beings that would render us morally responsible in the sense laid out in R2, it would not follow that we could not bear the sort of moral responsibility characterized in R1. This is not to say that we could, of course, but merely that whether we could would remain something of an open question. Furthermore, even if Strawson were correct in holding that R2 accurately describes the ordinary conception of moral responsibility, this would not settle the question of what it is to be genuinely morally responsible anymore than accurately characterizing the common conception of free will, for example, (if there is such a thing) would settle the question of whether free will requires alternative possibilities. Since it is not at all uncommon for people to have false beliefs about a wide range of ethical matters, one would need to provide some argument showing why we ought not to regard the ordinary folk conception of moral responsibility—to the extent that it is in fact the ordinary folk conception—as simply mistaken, at least in some respect (2005, p. 22).

Is it true that the conceptions of moral responsibility characterized in R1 and R2 are not equivalent but quite distinct, as Clarke contends? I suggest that they are not. R1 states that one who is truly morally responsible can be justly punished and rewarded for what one does. Assuming that one can be punished or rewarded fairly only if one deserves punishment or
reward, R1 also implies that one who is truly morally responsible can deserve punishment and reward for what one does. Importantly, however, it says nothing about what kind or amount of punishment and reward one who is responsible in this way can deserve. And rightly so. Whether one is (or can be) responsible in such a way that one can deserve punishment and reward for what one does is a question of moral responsibility. What type of punishment and reward one who is responsible in this way can deserve and so fairly receive, however, is (at least partly) a question of justice. (More on this later.) As a result, it is not necessarily the case that one who deserves an infinite amount of punishment is morally responsible for what one does in a way that one who deserves only a finite amount of punishment is not. Both could be equally morally responsible for what they do, even morally responsible in exactly the same way—namely, in such a way that they can deserve punishment and reward—and yet deserve different kinds or amounts of punishment. In this case, what explains the difference between what each of them deserves could be a difference in the wrong actions each has performed (or the harm each has caused) rather than a difference in the kind of responsibility each of them bears.

Although the fact that one deserves some amount of punishment does not entail that one does deserve an infinite amount of punishment, it does entail that, all else being equal, one could deserve an infinite amount of punishment. Why the qualification? While not logically entailed by R1, the possibility of deserving an infinite amount of punishment follows from the possibility of deserving any amount of punishment, given certain widely-recognized principles of justice. Indeed, the only case in which the possibility of deserving an infinite amount of punishment would not follow from the possibility of deserving some amount of punishment would be if there were some principle of justice to the effect that, regardless of what one does, it is impossible to deserve an infinite amount of punishment because such punishment is inherently unjust. Such a principle seems implausible. Consider, for instance, a person, Travis, who in order to feel better about himself stops by the local playground every day after work and pummels a few elementary school children—the smaller and weaker, the better. If Travis is morally responsible in such a way that can render some punishment just or fair, then it certainly seems that after his first trip to the playground Travis deserves some punishment for what he has done. Moreover, if Travis

---

66 It should be noted that although a person’s deserving punishment for what she does entails that it would be just or fair for her to be punished, this does not imply that punishing her is the right thing to do, nor even that it is morally permissible, all things considered. There may be cases in
deserves some amount of punishment for what he does after his first trip to the playground, then it would seem to follow that with each additional trip, with every child he pummels, Travis deserves an additional amount of punishment. Thus, if Travis were somehow able to pummel an infinite number of children, there is nothing that would prevent him from deserving an infinite amount of punishment (again, unless there were some principle of justice stating that under no conditions is an infinite amount of punishment ever just or fair, in which case Travis could theoretically reach a point where he could go on battering children and not deserve any additional punishment; but this seems absurd). Therefore, it looks as though it is no good arguing that Strawson’s argument, even if successful in showing to be impossible the sort of responsibility described in R2, fails to demonstrate that one could not be morally responsible in the sense suggested by R1. Since R1 and R2 mutually entail each other (again, assuming certain widely-recognized principles of justice), if Strawson succeeds in proving that it is impossible to be morally responsible in the one sense, he succeeds in doing so for both. But perhaps this is not the end of it.

5.3.3 The Real Problem with Heaven-and-Hell Responsibility

Although showing that it is impossible to bear the sort of responsibility specified by R2 would thereby also demonstrate the impossibility of being responsible in the way described in R1, it is not R2 moral responsibility that Strawson addresses. As Strawson makes very clear, it is heaven-and-hell responsibility that he takes to be the common conception and with which he is concerned. It is this sort of responsibility, he argues, that requires that one be causa sui, and it is responsibility of this sort that he takes to be impossible. The important thing to see, however, is that heaven-and-hell responsibility, as Strawson describes it, is characterized not by R2 but by

R3 True (or ultimate) responsibility is of such a kind that it can exist if and only

which, even though a person deserves punishment, it would be morally impermissible to punish her. This might be the case if, for example, circumstances were such that punishing a particular person who deserved to be punished would result in the harming of many innocent people. While this would not change the fact that this person still deserves punishment, and perhaps even that such punishment would be just or fair (at least to that individual), punishing her in this case would not, all things considered, be the morally correct thing to do.
if infinite punishment and reward conferred on finite human beings for the good and evil done in their lifetimes can be fair or just without having any pragmatic justification.

The mere possibility of fair or just infinite punishment and reward, though sufficient for the possibility of being morally responsible in the way spelled out in R2, is insufficient to establish the possibility of heaven-and-hell responsibility as Strawson describes it. Even supposing there are possible agents who, perhaps owing to their ability to commit an infinite amount of right and wrong actions, could deserve an infinite amount of punishment and reward, one thing seems clear: we are not they. According to Strawson, heaven-and-hell responsibility is of such a kind that it can exist if and only if “it could be just to punish some of us with (eternal) torment in hell and reward others with (eternal) bliss in heaven” (1994, p. 9; emphasis mine). Therefore, it is R3 that properly characterizes the conception of moral responsibility employed in Strawson’s argument.

Given that it is R3, and not R2, which accurately describes Strawson’s heaven-and-hell conception of moral responsibility, Strawson might well be correct in holding that it is impossible for anyone to be morally responsible in this way. What we should not fail to observe, however, is that the impossibility of our bearing the kind of responsibility identified by R3 has nothing to do with our inability to be self-caused. Instead, it is a function of the fact that human beings are capable of producing only a finite amount of good or evil during the course of their natural lives and so cannot possibly deserve an infinite amount of punishment or reward.

The heaven-and-hell conception of moral responsibility Strawson describes is a desert-entailing type of responsibility, one that could justify punishment and reward on purely retributivist grounds. But even though retributivism denies the need to appeal to pragmatic or utilitarian considerations in order to justify punishment and reward, it does not deny that punishment is appropriate only when proportional to the wrongdoing for which it is imposed. Indeed, that punishment is never justified when it is excessive in relation to the magnitude of the committed offense is something of a commonplace.

The problem with moral responsibility of the sort Strawson describes is that it is possible only if some human beings could be justly punished with eternal torment in hell. Such punishment will be just, however, only if it is proportional to the wrongdoing of those being
punished. Since we are capable of performing only finite actions and we have only a limited time in which to perform them, it is impossible for us to do anything that could make us deserving of the sort of punishment Strawson describes. As Kenneth Himma puts the point:

No matter how proficient at wrongdoing any person is, that wrongdoing will always be finite: since each instance of wrongdoing instantiates finite moral disvalue and no person can commit more than a finite number of instances of wrongdoing, the total quantity of wrongdoing accomplished by even the worst of persons will necessarily be finite. Thus…it would be unjust to punish someone who has lived only seventy or so years with eternal suffering because an infinite punishment is out of proportion to even cumulatively finite wrongdoing (2003, p. 64; note omitted).

Thus, even if one were to assume the strictest of retributivist standards—the principle of *lex talionis* (“an eye for eye”), for example—given the finite nature of human beings, there is nothing one could do, no amount of evil one could perform, nor harm one could cause, that would render infinite punishment proportional. Furthermore, as Marilyn McCord Adams (1999) has argued, it is not obvious that punishment of this sort could be proportional (and so, just) even if one were capable of causing an infinite amount of harm. She writes:

Suppose I knock out one tooth out of the mouth of each of thirty-two people each of whom has a full set of teeth and then the authorities knock out all thirty-two of my teeth by way of punishment. Is my having no teeth not much worse than their each having thirty-one teeth? Or suppose I interrupt television transmission of the superbowl game, thereby causing twenty million fans one hour of fury and frustration each. Surely, my suffering twenty million hours of fury and frustration is much worse (1999, p. 40).

The problem with the notion of punishing finite human beings with eternal torment in hell, then, is that it is inherently unjust. As Clark Pinnock puts it, “it is intolerable from a moral point of view because it makes God into a bloodthirsty monster who maintains an everlasting Auschwitz for victims whom he does not even allow to die!” (1990, p. 253)
It is because it would require that one could justly deserve disproportional punishment, which is inherently unjust, that it is impossible to be morally responsible in the way R3 describes. Causa sui considerations are irrelevant. Even if it were possible for us to be causa sui, and therefore completely responsible for the way we are, this would do nothing to change the fact that excessive punishment is inherently unjust, nor would it render us capable of infinite evil. Thus, the question of whether it is possible for us to be self-caused looks to be something of a red herring, at least as far as heaven-and-hell responsibility is concerned.

This is not to say that causa sui considerations have absolutely no bearing on questions of moral responsibility. Neither is it to say that one who is self-caused could not be morally responsible for what one does in a way that one who is not self-caused could not. On the contrary, it seems entirely plausible to suppose that, all else being equal, one who is self-caused would be more morally responsible for performing a particular action than one would have been had one not been self-caused. Importantly, however, the difference in responsibility in these two cases would be a matter of degree, not type. In this way, it would be like the difference between the responsibility borne by children and that borne by adults. Returning to our earlier case, suppose that instead of Travis, it is Travis’s twelve-year-old son, Joe, who likes to spend his afternoons pummeling the younger children on the playground. Assuming that Joe is an ordinary twelve-year-old boy, it certainly seems that he is genuinely morally responsible (indeed, blameworthy) for his actions. Furthermore, his responsibility appears to be of such a kind that he deserves some punishment for what he has done. What seems equally obvious, however, is that Joe is not as morally responsible, nor is he deserving of as much punishment, as Travis would be for engaging in the very same behavior, even assuming that the harm caused in both cases is the same. Here, the fact that Joe and Travis each deserve different amounts of punishment does not show that each bears a different kind of moral responsibility. Both of them are responsible in the same sort of way, namely, in a way that renders them capable of deserving punishment for what they do. It is just that, owing to his greater moral knowledge and ability to control himself (as well, perhaps, as other related factors), Travis is morally responsible to a greater degree. Similarly, since what one does is at least partly (though perhaps not entirely) a function of the way one is, it is plausible to suppose that one who is causa sui, and so completely responsible for the way one is, would be more morally responsible for what one does than one whose actions are at least partly a function of factors for which one is not completely (or at all) responsible (e.g.,
one’s genetic predispositions and childhood environment). Here, too, though, the difference is merely a matter of degree. Moreover, although being *causa sui* might make it possible for one to be more morally responsible (i.e., responsible to a greater degree) for what one does than one could otherwise be, it would not make it possible to be morally responsible in the way R3 describes.

### 5.3.4 Why This Matters

I have been at pains to show that *causa sui* considerations are irrelevant to the fact that the conception of moral responsibility specified in R3 is impossible, for even if it were possible for one to be *causa sui* this would not affect the impossibility of being morally responsible in the sense laid out in R3. But why does this matter? It is not merely the fact that *causa sui* considerations are irrelevant to the impossibility of R3 moral responsibility but the reason they are irrelevant that is so important, namely, R3’s requirement that it could be fair or just to punish finite human beings with eternal torment in hell. Since human beings are capable only of finite wrongdoing, infinite punishment in hell will always be disproportional to the wrong one has committed. In effect, then, R3 states that one can be truly morally responsible for what one does only if punishment that is inherently unjust can be just. It is this feature of R3 that makes responsibility of the heaven-and-hell variety impossible, and it is this feature of R3 that makes the possession of alternative possibilities irrelevant to whether we are capability of bearing this type of moral responsibility. Most importantly, it is a feature not shared by R1. This is significant because, since R1 is a thesis only about moral responsibility and says nothing about what type of punishment and reward it can fair to confer on finite human beings, being responsible in the way set out in R1 does not entail being responsible in the way set out in R3. Thus, even if the sort of responsibility identified in R3 is impossible, this does not show that it is impossible to be morally responsible in the way described in R1, namely, in a way that can render us capable of deserving at least some amount punishment and reward. Similarly, even if the ability to do otherwise is irrelevant to the sort of responsibility identified in R3, it may well be that alternative possibilities are required for the kind of responsibility characterized by R1.

Furthermore, as Clarke points out, even if Strawson were correct in claiming that heaven-and-hell responsibility is the conception of responsibility held by most people, this would not settle the question of what genuine responsibility is. In fact, to the extent that our intuition that
we are indeed genuinely morally responsible for at least some of what we do is stronger or more fundamental than whatever intuitions we might have concerning the precise nature of that responsibility, in showing that it is impossible to be morally responsible in the sense specified by R3 Strawson might very well be providing us with good reason to reject R3 in favor of some other conception of moral responsibility—one that is possible for persons such as ourselves to satisfy.

Here it is perhaps worth noting that there appears to be at least one good reason for thinking that R3, heaven-and-hell responsibility, is not, contra Strawson, the common conception of moral responsibility, which is this: R3 is a thesis about more than just moral responsibility and desert. Earlier I stated that whereas whether and when one is (or can be) morally responsible and deserving of punishment and reward is a question of responsibility, what type of punishment and reward it can be fair to confer for certain kinds of action (e.g., actions that produce only a finite amount of good or harm) is a question of justice. Insofar as R3 offers (at least partial) answers to both of these questions, it is a thesis not only about the nature of moral responsibility but also about the nature of justice and fairness. Moreover, it is a thesis that relates these issues in such a way that holding it would require that one accept, in addition to its separate parts, that each part is true if and only if the other part is true. It is for this reason that I think it unlikely that R3 describes the common conception of moral responsibility. It is entirely plausible that many, if not most, people take themselves to be truly morally responsible for at least some of what they do (even in a way that can make it fair or just to punish and reward them for their actions). It might well also be true (though this is more controversial) that many, perhaps even most, people believe that it could be just to punish some of us with eternal torment in hell and reward others of us with eternal bliss in heaven. What strikes me as utterly implausible, however, is the supposition that most, or even many, people believe that they are truly morally responsible for what they do if and only if it could be fair or just to punish some of us with eternal torment in hell and reward others of us with eternal bliss in heaven.

Those who hold R3 as their conception of moral responsibility would have to say that were they convinced that no human beings, no matter what they do, could deserve to be punished with eternal torment in hell, they would thereby be forced to give up altogether the idea that people are truly morally responsible for what they do. But just as there is no reason to think that people who disagree about, say, whether capital punishment or castration is ever justified must
also disagree about what it is to be genuinely morally responsible for what one does, there seems to be no reason to think that those who disagree about whether eternal punishment in hell could ever be just or fair must therefore be operating under different conceptions of moral responsibility. It is not the nature of responsibility about which such people disagree but, rather, the nature of justice and what types of punishment are fair. Since R3 entails that those who disagree about whether anyone could deserve to be punished with eternal torment in hell must also disagree about what it is to be genuinely morally responsible, it seems to me highly unlikely that R3 is the ordinary conception of moral responsibility.

My aim in this section has been to give careful consideration to the particular conception of responsibility doing the work in Strawson’s argument for the impossibility of moral responsibility, particularly since thinking about responsibility in this way straightforwardly leads to the view that whether we are morally responsible for things we do in no way depends on whether we are able to do otherwise. Contra Clarke, I have argued that the kind of moral responsibility characterized in R1 is in fact equivalent to that characterized in R2. Since they describe equivalent conceptions of moral responsibility, demonstrating the impossibility of the kind of responsibility specified in R2 would also show that it is impossible to be responsible in the way suggested by R1. But heaven-and-hell responsibility as Strawson describes it is accurately characterized not by R2 but by R3. R3, however, is a thesis about more than just moral responsibility and desert; it is also a thesis about justice and fairness. In particular, it is about the type and amount of punishment and reward that it can be fair to confer on finite human beings for the actions they perform in their lifetimes. It is this feature of R3 which makes responsibility of the heaven-and-hell variety impossible. Moreover, it is a feature not shared by R1. This is significant because, since R1 is a thesis only about moral responsibility and says nothing about what type of punishment and reward it can fair to confer on finite human beings, being responsible in the way set out in R1 does not entail being responsible in the way set out in R3. It may well be (indeed, I suspect) that in treating R1 and R3 as equivalent conceptions of true (or ultimate) moral responsibility Strawson has allowed this further feature of R3 to do work in his attempt to demonstrate the impossibility of R1—work which would be illicit since it is not a feature R1 shares. At any rate, given that they are in fact not equivalent conceptions, even if the sort of responsibility identified in R3 is impossible this does not show that it is impossible to be morally responsible in the way described in R1, namely, in a way that would render us capable
of deserving praise and blame, reward and punishment, for what we do. In the same way, even if alternative possibilities are irrelevant to whether we can be heaven-and-hell responsible for the things that we do, they might well be crucial to whether our actions are ever things for which we can deserve (finite) praise or blame, punishment or reward.

Where does this leave Strawson’s argument? It certainly remains open to him to simply dispense with R3 and then run the same basic argument against ordinary desert-entailing responsibility of the sort found in R1 (indeed, this is precisely what I think he should do), but this in itself is not inconsequential. Though a concession on this point does not make the argument go away, neither does it allow it to proceed exactly as before. Once divorced from a conception of responsibility of the sort characterized in R3 (heaven-and-hell responsibility)—the intuitive impossibility of which seems to have distracted from the reason for its impossibility—the argument loses rhetorical force. This has the further result of drawing more clearly into focus, and placing additional pressure on, some of the more controversial points in Strawson’s argument that are left to do the work, to which there is no shortage of objections in the literature.\footnote{See, for example, Clarke 1997 and 2003 (pp. 170-176), Honderich 1988 (pp.178-80), Hurley 2000, Mele 1995 (pp. 221-227), and Pereboom 2001 (pp. 65-68).} Strawson’s basic argument will be compelling, then, only to the extent that the crucial moves are adequately supported and the objections answered. Whether such a defense is forthcoming, and how successful it will be, remains to be seen. In this task, however, the repugnance of the conception of justice and fairness embedded in R3 should not be allowed to do any of the heavy lifting.

5.4 Conclusion

My main goal in this chapter has been to draw attention to recent attempts to characterize moral responsibility in a way that differs from the traditional desert-entailing account of moral responsibility. It is not my purpose to deny that there might be different kinds (or senses) of moral responsibility that agents can bear or to insist that desert-entailing accounts of moral responsibility alone are deserving of that name (though I lean strongly in that direction). What I wish to combat, however, is any use of these distinctions to settle the longstanding debates surrounding free will, determinism, moral responsibility, and the ability to do otherwise—
especially when these distinctions are not being made explicit. Even if we allow, for instance, that some of the deflationary accounts that have been considered describe genuine kinds of moral responsibility and so allow that there are genuine kinds of moral responsibility that do not require alternative possibilities, these are not the kind of moral responsibility that is at issue in PAP. The reason for this is not merely that these accounts imply that moral responsibility does not require the ability to do otherwise (or the falsity of determinism) but that they so obviously imply this. Consequently, if these were adequate characterizations of the moral responsibility at issue in PAP (or the compatibility dispute), not only would Frankfurt-style cases be wholly unnecessary to show that PAP is false, the entire debate is simply wrongheaded and obviously so. Rather than show this debate to be utterly mysterious, however, what this indicates is that those advocating non-desert-entailing kinds of moral responsibility are not talking about the same thing that proponents of PAP are talking about. Thus, even if, as seems so clear as to hardly require argument, being morally responsible for what one has done in the senses described by these alternative conceptions of moral responsibility does not require that one could have done otherwise, this fails to provide us with any reason to reject PAP.
CHAPTER SIX

CONCLUSION

My aim in this dissertation has been to defend PAP against two major lines of attack: Frankfurt-style cases and certain revisionist conceptions of moral responsibility. I have argued that Frankfurt-style cases fail to provide us with good reason to reject the view that moral responsibility requires alternative possibilities. Although these cases do not show PAP to be false, they are helpful in teaching us how this principle ought to be precisely formulated and understood. They are also useful in helping us to identify the bare minimum alternative possibilities that enable an agent to be morally responsible for what she does. I have also argued that the debate surrounding PAP will not (and should not) be settled merely by changing the way we apply the term ‘morally responsible’. Whether a person can be morally responsible for what he has done only if he could have done otherwise is a substantive issue, not merely a semantic one, and showing that PAP is false will require substantive reasons.
REFERENCES


BIOGRAPHICAL SKETCH

Michael Robinson was born on February 18, 1981, in Santa Monica, California. He holds a B.A. in Philosophy from Biola University, an M.A. in Philosophy from Texas Tech University, and a Ph.D. in Philosophy from Florida State University. His primary research interests lie at the intersection of Ethics and Philosophy of Action, particularly with issues concerning free will and moral responsibility.