FREEDOM, ACTION, AND THE CONDITIONS FOR BLAME AND PRAISE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Matthew Ryan Paskell

May 2022

FREEDOM, ACTION, AND THE CONDITIONS FOR BLAME AND PRAISE

Matthew Ryan Paskell, Ph.D.

Cornell University 2022

This work concerns the nature of moral responsibility, and in particular the conditions under which blame and praise may be appropriate. Each chapter focuses on a different kind of condition for moral responsibility—responsiveness to reasons, the ability to do otherwise, or having the right motivation for action. In Chapter 1, I argue that reasons-responsiveness is neither necessary nor explanatory as a control condition for moral responsibility. I present a counterexample to this proposed condition involving an agent who's willfully impulsive and show that responsiveness to reasons cannot be the explanation for his blameworthiness. Instead, I suggest that blameworthiness is explained by his ability to *settle* on a particular action, which is a condition that can be met with or without satisfying reasons-responsiveness conditions. In Chapter 2, I defend the "Principle of Alternative Possibilities" (PAP) against a certain kind of Frankfurt-style case—buffered alternatives. I develop a new dilemma for buffered alternatives cases which shows that they fail to constitute counterexamples to the PAP. I argue that the introduction of the buffer, though it may remove certain alternative possibilities, shifts the locus of responsibility to earlier provisional or character-forming decisions. In Chapter 3, I assess accounts of positive moral worth that take *doing the right thing for the right reason* to be necessary, and perhaps sufficient, for praiseworthiness. I evaluate these accounts by applying them to several challenging cases, with the goal of determining whether they accord with intuitive judgments. This involves varying either the nature of the action itself or the motivations for action and then determining what the accounts would and ought to say about praiseworthiness. While I conclude that the accounts do well at explaining our intuitions about difficult cases, I argue that some *right thing for the right reason* accounts have advantages over others in certain contexts.

BIOGRAPHICAL SKETCH

Matt grew up in Sierra Vista, Arizona, the middle of five children. In 2007 he narrowly graduated from high school, after which he worked various menial jobs to make ends meet. During that time, he became interested in life's biggest questions from both a philosophical and scientific perspective. In 2010 he moved to Tucson, Arizona and enrolled at Pima Community College, becoming the first in his family to attend college. There he focused his studies on Philosophy and Astronomy, graduating in 2012 with Highest Honors. He received an Associate of Arts in Liberal Arts as well as a scholarship to attend the University of Arizona. While at the University of Arizona, Matt majored in Philosophy and minored in Liberal Arts Astronomy. He wrote an honors thesis on the topic of free will and moral responsibility, titled "Manipulation, Argument, and Experiment: Putting Folk Intuitions into Context." In 2014 he graduated *summa cum laude* with a Bachelor of Arts in Philosophy, while also receiving the department's "Undergraduate Riesen Prize" for outstanding academic achievement. After completing his undergraduate studies, Matt moved to Ithaca, New York to attend Cornell University as a graduate student in Philosophy. He continued his work on free will and moral responsibility, receiving a Master of Arts in 2017 and completing his Ph.D. in 2022.

For those condemned to be free and seeking relief.

ACKNOWLEDGMENTS

While reflecting on my path to completing this dissertation, three people in particular stand out to me. First, I want to thank Kent Slinker for helping me to recognize my passion for philosophy at a time when I didn't quite know what that meant. Next, I want to thank Michael McKenna both for helping me to cultivate my philosophical understanding and ensuring that I was in the best position to continue my studies in philosophy. Finally, I want to thank Derk Pereboom for helping me to refine my philosophical abilities, to the point where I'm no longer merely doing philosophy, but can now also confidently call myself *a philosopher*.

In addition, I'd like to thank Julia Markovits, Shaun Nichols, Pam Hannah, Dorothy Vanderbilt, and the many other faculty and staff members in Cornell's philosophy department who provided feedback, encouragement, and support over the years. I also owe immense thanks to the fellow graduate students, friends, and others who made my time in Ithaca worthwhile.

TABLE OF CONTENTS

CHAPTER 1

RESPONSIBILITY AND IMPULSIVE MECHANISMS

1.1 Introduction

One of the most common strains of compatibilism in the contemporary literature on moral responsibility focuses on reasons-responsiveness. These accounts take the primary control condition for responsibility to be an appropriate sensitivity to reasons, with the characterization of "appropriateness" varying among theorists.[1] The most prominent view of this type was developed by John Martin Fischer and Mark Ravizza (1998) and defended by Fischer in a number of further works.[2] On their view appropriate sensitivity to reasons involves being *moderately* reasons-responsive, which is taken to be a necessary condition for responsibility. Although intuitively responsible agents typically satisfy this condition, I'll argue that consideration of a certain type of case—freely chosen impulsivity—gives good reason to think that moderate reasons-responsiveness is not in fact necessary for responsibility.[3] Moreover, I intend to demonstrate that this type of compatibilism doesn't do the explanatory work that a theory of responsibility should. The discussion will be framed by the paradigm account

[1] The type of moral responsibility typically at issue in this debate, and the type that I'll be concerned with, is basic desert responsibility. We can take this to mean that an agent who is basically responsible for an action is responsible, and would deserve praise or blame, solely in virtue of having knowingly performed a wrong action and not in virtue of any consequentialist or contractualist considerations. I'll use the term 'responsible' throughout as shorthand for this type of moral responsibility.

[2] See, e.g., Fischer (2006), (2009), and (2012).

[3] As I hope will become clear, I don't intend the term 'impulsive' to indicate a lack of control, as it often does in ordinary usage. Rather, I'll use the term to emphasize that it's an agent's *immediate,* and perhaps *strongest* desire.

set out by Fischer and Ravizza, although my conclusions will extend to other reasons-responsiveness theories as well.[4] After outlining a framework for reasons-responsiveness theories, I'll lay out the challenge and explore potential responses on behalf of Fischer and Ravizza. Finally, I'll offer my own diagnosis of the cases at hand.

1.2 Fischer and Ravizza's Reasons-Responsiveness

The theory Fischer and Ravizza develop is an *actual sequence* view, as they maintain that the actual history of an action is what's relevant to responsibility and not genuine, metaphysical access to alternative possibilities.[5] The actual psychological history that results in an action is referred to as the agent's mechanism, and two mechanism-relevant requirements must be met for an agent to be responsible. First, in order to preclude certain manipulated agents from being responsible, the mechanism must be "*the agent's* own" (Fischer, 2012a, p.187). This ownership condition has three components:

> First, an individual must see himself as the source of his behavior…in the sense that he must see that his choices and actions are efficacious in the world…Second, the individual must accept that he is a fair target of the reactive attitudes as a result of how he exercises this agency in certain contexts…The third condition on taking responsibility requires that the individual's view of himself specified in the first two conditions be based, in an appropriate way, on the evidence. (Fischer 2012a: p.190)[6]

The other, core requirement for responsibility is that the mechanism from which the action results must be appropriately reasons responsive. This reasons-responsiveness

[4] See, e.g., Nelkin (2011), McKenna (2013), and Sartorio (2016).
[5] This is not to say that certain counterfactual claims aren't relevant at all, since these will help determine "modal or dispositional properties of the actual sequence" (Fischer, 2012b, p.124).
[6] Fischer is open to eliminating the second condition in light of certain cases (see, e.g., Mele (2006a)).

condition could potentially be characterized in number of ways. In a very strong form, reasons-responsiveness would require that anytime there is a sufficient reason to do otherwise the agent recognizes and acts on that reason. A very weak reasons-responsiveness condition would require that there be at least one counterfactual scenario in which the agent recognized and reacted to a sufficient reason. As Fischer acknowledges, neither of *these* forms will be appropriate (Fischer 2012b: p.125). This is because strong reasons-responsiveness fails to preserve responsibility in weakness of will cases while weak reasons-responsiveness would entail that those with no sane pattern of responsiveness,[7] or debilitating psychological conditions, would be incorrectly deemed responsible in many cases. Instead, Fischer and Ravizza opt for a *moderate* reasons-responsiveness condition—given an agent *S* whose action *X* issues from a mechanism *K*:

> "*K* is moderately reasons-responsive if and only if there is a range of possible scenarios *R* in which a *K*-type mechanism operates such that: (i) *S* recognizes in *R* what can be seen from an appropriate third-party perspective as an understandable pattern of sufficient reasons for not doing *X*; and (ii) there is at least one such scenario in which *S* refrains from doing [*X*] for such a reason." (Fischer, 2012b, p.125)

As comes out in this characterization, there are two components to Fischer-Ravizza style reasons-responsiveness. First, there's the condition that the agent is in some sense moderately *receptive* to reasons for doing otherwise (i). Next, there's a very weak

[7] This particular problem for a weak reasons-responsiveness condition becomes clearer through example. Imagine that David the car salesman will show up for work every day unless the temperature outside is 100°F. If extreme heat were the reason that he would miss work then it seems he could be appropriately responsive in the responsibility-conferring manner that Fischer and Ravizza have in mind. However, if we suppose that David *would still* show up to work if the temperature were 105°F, or that he would only show up to work if the temperature were a prime number, we get a different, incorrect result. Even though there is a counterfactual scenario where he recognizes and acts on a reason for doing other than he actually does it seems that the pattern of reasons is responsibility-undermining. Other cases of this sort are discussed in Fischer & Ravizza (1998, pp.65-8).

reactivity condition, requiring that there be just one counterfactual scenario in which the agent acts on a reason to do otherwise (ii).[8] In what follows I'll understand receptivity as involving recognition and evaluation of reasons and reactivity as "the capacity to *translate* reasons into choices" (Fischer and Ravizza, 1998, p.69; Fischer, 2012a, p.188).

1.3 The Case of George

Having outlined the main components of Fischer and Ravizza's version of reasons-responsiveness, I'll now set up a challenge for the account. Given a theory of moral responsibility that requires agents to be reasons-responsive, one question that arises is whether our responsiveness to reasons is itself something we have control over. It's obvious that we sometimes choose to ignore reasons for alternative actions, or choose not to act in accordance with reasons we recognize, and behave poorly as a result. After all, agents are often considered blameworthy in just these sorts of cases and reasons-responsiveness accounts have no trouble diagnosing them. More difficult cases, however, involve voluntarily choosing not to respond to reasons in the long term. Evaluation of this sort of agent, who *decides* not to be receptive to reasons to act otherwise, is not merely a difficulty for Fischer and Ravizza. As I'll argue, tracing the agent's responsibility for later transgressions back to the time of that initial decision is

[8] Fischer has since acknowledged that the reactivity condition as originally formulated may be too weak and has indicated that he might accept a shift to regular receptivity and *weaker* reactivity (2005, 2012a). This admission comes in light of counterexamples outlined by Mele (2000) and McKenna (2005). My challenge to reasons-responsiveness will focus primarily on the receptivity component of the theory and so this potential revision will not impact the discussion.

inappropriate in the cases I present. Thus, the agent's responsibility for such transgressions reveals that reasons-responsiveness is not a necessary condition for responsibility at all, nor does an agent's responsiveness to reasons *explain* why they are intuitively responsible in particular cases. I begin with a case involving an agent named George:

> **George**
> George is an ordinary human agent who has been moderately reasons-responsive his entire adult life. One New Year's Eve, however, concerned that his overthinking is complicating his life, George makes a resolution to be more impulsive. He vows that in the coming year he will act on all of his immediate desires unless doing so would involve an imminent threat to his life. As the year progresses George sticks to his resolution—occasionally acting laudably (e.g., visiting his mother in the hospital) and occasionally acting objectionably (e.g., illicitly parking in a handicapped space), all the while paying no attention to reasons to do otherwise. He sees himself, perhaps now more than ever, as the source of his actions. He also realizes that some of his actions have moral significance, and that when others censure him they do so appropriately. On the first day of May, George orders take-out at a restaurant and, after receiving his food, grabs money out of the employees' tip jar and leaves.[9]

What should we say about George's responsibility in this case? First, he clearly satisfies the ownership conditions outlined by Fischer and Ravizza, both prior to and at the time of his theft. There is no reason to think that acting impulsively in this way would reduce his feeling of sourcehood, and we can assume that he has no evidence that this feeling is illusory. George can fully identify with his desires and actions and his phenomenological experience—from decision to action to outcome—need not be any different in kind than that of any ordinary agent. This includes viewing the effects of his actions as flowing *from him*. It's also easy to suppose that he sees others, and himself,

[9] One thing to note is that this case differs from the type of "nonreflective behavior" Fischer and Ravizza consider and adequately address (Fischer & Ravizza, 1998, pp.85-9). Rather than acting out of habit, or coasting on autopilot, we can think of George as being struck with impulses to perform certain actions and always assenting to those impulses. That is, the path from desire to action need not be "automatic."

as a fair target of the reactive attitudes. The case could even be specified further so that a third-person observer would judge that George sees himself this way—he may occasionally condemn others when they act distastefully, and he might, if he has the immediate desire, apologize after committing a wrong. Ownership can be satisfied independently of other conditions.

The other mechanism-related requirement, that the action results from a mechanism that is moderately-reasons responsive, requires more discussion. George is certainly *reactive* to reasons given Fischer and Ravizza's lenient characterization. As stipulated, he would not steal if a threat to his life were presented, and so there is at least one counterfactual scenario in which he would have acted other than he did. The issue is that, although he retains the capacity for being appropriately receptive, he chooses not to engage this capacity. As I hope will become clear in the following section, my challenge doesn't rest on the claim that George is not moderately reasons responsive. In fact, I believe that he *is* reasons responsive given standard characterizations of the conditions, including those of Fischer and Ravizza. What I intend to show is that, despite of this, comparing George to other, similar agents reveals that reasons-responsiveness is neither necessary nor explanatory when evaluating George's responsibility.

1.4 The Challenge: George and His Counterparts

Given Fischer and Ravizza's characterization of moderate reasons-responsiveness, George would be considered directly responsible for his theft.[10] This is the case even

[10] By "directly responsible" I simply mean non-derivatively responsible. That is, responsible for the action simpliciter and not in virtue of some other, prior, action.

though at the time he steals he's not *actually* receptive to reasons to do otherwise—he simply sees the money in the jar and takes it.[11] The receptivity component of reasons-responsiveness involves the *capacity* to recognize and evaluate reasons for action and this can be demonstrated *either* by considering counterfactual possibilities *or* actual behavior. What's important for my purposes is that George meets the criteria for reasons-responsiveness in virtue of the former and not the latter. Even if we assume that George recognizes reasons in the appropriate way, he is not actually evaluating them. Upon seeing a security camera in the restaurant, he may recognize a reason not to steal but he doesn't weigh it against any other reasons. We can suppose that there is no evaluation going on at all—George simply notes that reason and acts on his desire for the money. Moreover, it's not clear that we even have to suppose that George recognizes reasons *in the appropriate way*. Most people who feel an impulse to take the money in such a situation, and who see the security camera, would recognize *a reason not to steal*. George may see the security camera, thereby recognizing a reason not to steal, but not recognize it as something that should factor into his decision to steal. He recognizes the thing that *is* a reason to curb immoral behavior, but he doesn't recognize it *as* a thing in light of which *he* ought to consider modifying his behavior. If George is responsible for stealing, then, it's not because he is *actually* responding to reasons at the time of his action.

The reason George would be considered directly responsible on Fischer and Ravizza's view is that, even though he may not actually be responding to reasons at the

[11] This is not to say that George isn't acting on reasons at all. There is one conspicuous reason he's acting for—that "this" is his immediate desire.

time of the theft, he nevertheless retains the *capacity* to recognize and evaluate reasons for and against performing that action. In general, the requirement that one only have the capacity to be appropriately reasons-responsive is well-motivated. Those who are blameworthy on reasons-responsiveness accounts may, for whatever reason, not actually respond to reasons in the process of performing some bad action, but the fact that they *could have* is enough for holding responsible. For example, one who touches the artwork in the museum while consciously ignoring the prohibiting signs may still be responsible for any unwelcome consequences, so long as the relevant capacities are still present.

However, while it would be consistent with the theory to say that George is directly responsible because he retained the capacity for appropriate receptivity, this does not seem to be doing the explanatory work in this case. For George, recognizing and evaluating reasons for modifying his behavior is just an idle tool that he has no intention to use, neither at the time of his action nor anytime in the near future. The mere fact that it was available to him does not appear relevant to his responsibility in the right kind of way. That retaining the capacity for reasons-responsiveness is irrelevant to whether George is responsible can be illustrated by comparison to another case. I'll present two versions of this case, each with certain advantages. I begin with Gene, who we might think of as George (or his counterpart) in another possible world:

> **Gene**
> From birth, Gene is exactly like George in every respect, including all of his psychological processes and resulting actions. But the night before Gene goes to the take-out restaurant neuroscientists sneak into his apartment and inject him with a drug that disables his capacity for appropriate reasons-receptivity. For 24 hours thereafter, Gene no longer has the capacity to appropriately recognize and evaluate reasons. On the first day of May, Gene orders take-out at a restaurant

> and, after receiving his food, grabs money out of the employees' tip jar and leaves. Because Gene never attempts to evaluate reasons for an alternative action, the causal history of the action is identical to that of George.

The only difference between the "George" and "Gene" cases is that George retains the capacity to respond to reasons in the appropriate way while Gene, at the time of the theft, lacks the capacity. Neither George nor Gene actually attempts to evaluate reasons for doing otherwise. The neuroscientists' removing Gene's ability has no effect on how he actually acts—the psychological processes before, after, and during the theft are the same as they would have been without the intrusion and, of course, the same as George's. The presence of the neuroscientists does make it the case that George and Gene differ with respect to certain modal properties—while it's possible for George to consider reasons for alternative actions, the same cannot be said of Gene. This would not appear to be a responsibility-relevant difference, however, since neither actually attempt to consider any reasons. It's only a matter of luck that Gene could not have engaged this capacity had he tried. Is it really plausible, then, that George is directly responsible for his theft while Gene is not? Those unperturbed by this sort of luck playing a role in responsibility might not be troubled by this result, although I suspect that most would find a difference in responsibility implausible.[12] If George is directly responsible then so too is Gene, and an appeal to the capacity to appropriately respond to reasons won't explain this.

[12] Embracing an important element of Sartorio's reasons-responsiveness view would lend additional support for this conclusion. She argues that a certain supervenience claim, "*no difference in freedom without a difference in the relevant elements of the causal sequence,"* holds for the freedom necessary for moral responsibility (Sartorio, 2012, p.32). Applied to the George and Gene cases, there is no difference in the causal sequences leading to action, or at least no *relevant* difference, and so there is no difference in freedom (and thus no difference in responsibility).

A concern for this Gene case, and its being a suitable comparison for the George case, involves the issue of *sameness of mechanism*. Crucial to my assessment of their respective responsibility is that the causal history of the actions is the same in both cases. However, accepting Fischer and Ravizza's characterization of a mechanism, and their claim that we have "intuitions about fairly clear cases" regarding same and different mechanisms (Fischer and Ravizza, 1998, p.40), it seems plausible that both George and Gene's actions indeed issue from the same mechanism.[13] By the term 'mechanism,' Fischer and Ravizza aim to pick out "(nothing) over and above the process that leads to the relevant upshot," and maintain that we could instead invoke "the process that leads to the action, or the 'way the action comes about'" (Fischer and Ravizza, 1998, p.38). Intuitively, an unproductive injection does not change the relevant process. Still, the fact that there is *something* different in the George and Gene cases—the neuroscientists' actively disabling of a capacity in the latter—leaves room for the worry that this difference results in a change in the mechanism.

Given that there is no process of reasoning that George undertakes and Gene does not, their desires and motivations are the same from birth through the theft, and the causal story of how the action came about would be the same in both cases, a "difference in mechanism" response appears implausible. It is at least intuitive that both thefts issued from the same mechanism. But even if the intervention in Gene's case does make it so that the mechanisms issuing in action really are different, this still may not be enough to establish that the comparison is flawed. A plausible principle may be that, in

[13] "Same mechanism" can be understood as qualitatively identical, especially if there's concern about transworld identification.

order for a difference in mechanism to result in different judgements of blameworthiness, it would have to be the case that the difference is *responsibility-relevant*.[14] While a lack of capacity for reasons-responsiveness may often appear responsibility-relevant, in Gene's case the causal process leading to the action (including the action itself) would be the same with or without intervention. The same, indeed, as the processes leading to George's action. Of course, there is one conspicuous justification for a "difference in mechanism" response—that, unlike George, Gene may not be able to perform an alternative action given that he is unable to evaluate reasons for doing otherwise, and that this difference in the respective mechanisms is responsibility-relevant. However, this move isn't available to Fischer and Ravizza (or most other reasons-responsiveness theorists) because it requires accepting the "Principle of Alternative Possibilities" (PAP). This principle, roughly, holds that an agent can only be responsible if she could have done otherwise. Fischer and Ravizza's can't appeal to such a principle because access to alternative courses of action is not responsibility-relevant on actual sequence views.[15]

For those who remain unconvinced that the sort of intervention described in the Gene case preserves sameness of mechanism in a responsibility-relevant way, a different case may be more plausible. Here is another case involving Gene, although this time the intervention scenario is in the style that Harry Frankfurt originated as a challenge to the PAP (1969):

[14] Here we must hold fixed all other context-relevant features of the actions, of course.
[15] In particular, Fischer and Ravizza distinguish "regulative" control (which requires access to alternative possibilities) from "guidance" control (which requires no such access). They maintain that guidance control, and not regulative control, "grounds moral responsibility" (Fischer & Ravizza, 1998, pp.30-4).

Frankfurted Gene

Gene is exactly like George in every respect, including all of his psychological processes and resulting actions. But the night before Gene goes to the take-out restaurant, neuroengineers sneak into the restaurant and install both thought-reading and receptivity disabling technology in the ceiling. While it's possible that Gene could break his commitment to be impulsive, a necessary condition of such a reversion is that he first imagines his good friend Feldman chastising him for making his resolution in the first place. Upon thinking about Feldman, he could then decide to become receptive to reasons and refrain from stealing despite his initial desire to do so. If the neuroengineers detect that Gene is thinking of Feldman, however, they will disable his capacity to appropriately respond to reasons, thereby ensuring that he keeps his resolution and, ultimately, steals. As it happens, Gene does not think of Feldman, the neuroengineers do not intervene, and he steals just as George does. The mechanism from which this action issued is identical to that of George, and so too is the causal history of the action.[16]

Here again we have a case in which an agent, this time Frankfurted Gene (^{F}Gene), performs the same action as George and is intuitively just as blameworthy despite lacking the capacity to be appropriately reasons-responsive at the time of action.[17] The advantage of this case is that, unlike with Gene, the potential interveners do *nothing at all*. There should not be any concern, then, that the action of ^{F}Gene issued from a mechanism that differs in any way from George's.

At this point one might object that the mere presence of the neuroengineers is not enough to eliminate the capacity for appropriate reasons-responsiveness. This objection is pertinent since my claim is that George retains the capacity, ^{F}Gene (and Gene) do not, and yet assessments of responsibility should be the same in all cases. Judgments on this matter will likely mirror judgments about standard Frankfurt-style cases, with those who

[16] This Frankfurt-style case is modeled on the "buffered-alternatives" style, developed by Derk Pereboom (2001; 2014) and David Hunt (2005).

[17] When comparing the George and ^{F}Gene cases it should also be assumed that the same necessary condition is required for George to abandon his resolution (though without the prospect of counterfactual intervention).

believe that the ability to do otherwise is eliminated in standard Frankfurt-style cases also believing that ^{F}Gene's capacity to be receptive to reasons is eliminated.[18] The structures of these cases are exactly the same and the potential interveners merely seek to disable a different type of ability. Accordingly, ^{F}Gene's case may only be persuasive to those already convinced by Frankfurt and his defenders. But, as it happens, Fischer and Ravizza are among those persuaded by Frankfurt cases of one form or another (Fischer and Ravizza, 1998, pp.29-41).[19] Moreover, the majority of reasons-responsiveness theorists are also persuaded by Frankfurt-style cases, including McKenna (2013) and Sartorio (2016).[20]

As it stands, the "Gene" case ensures that the relevant capacity is eliminated but may allow room for disagreement over whether the mechanism issuing in his action is the same as George's. For ^{F}Gene, the mechanism is clearly the same as George's but some, perhaps those unconvinced by Frankfurt-style cases, may not agree that the mere presence of the neuroengineers eliminates the relevant capacity. While I favor the "Gene" case, at the very least "^{F}Gene" should resonate with source theorists and those opposed to "Gene" owe an explanation for the claim that there is a responsibility-relevant difference between the mechanisms of George and Gene.

[18] The ^{F}Gene case can also be restructured to match one's preferred style of Frankfurt case—for example, as a traditional Frankfurt case or as a trumping preemption Frankfurt-style case (e.g., Mele and Robb (1998))—rather than a buffered alternatives case.

[19] Fischer also has his own version of a Frankfurt case, contending that robust alternative possibilities are not necessary for responsibility (2004).

[20] A notable exception is Nelkin (2011). On her asymmetric, rational abilities view praiseworthiness does not require the ability to do otherwise. While blameworthiness *does* require the ability to do otherwise, she argues that Frankfurt-style cases don't eliminate the kind of ability relevant to her theory (Nelkin, 2011, p.115).

1.5 Alternative Assessments of Responsibility?

Although Fischer-Ravizza's theory would entail that George is directly responsible for his theft in virtue of his retaining the capacity for reasons-responsiveness, one might suggest instead that he is merely derivatively responsible. [21] With this in mind we may set aside this implication of the theory and consider what would be involved in this judgment. On this line of response, George is responsible for the theft, and for any other impermissible acts he may have committed throughout the year, only in virtue of his having freely resolved to be impulsive. It could be argued that this is what drives the intuition that he is indeed responsible. Although George meets all of the conditions Fischer and Ravizza require for direct responsibility, and some will have the clear intuition that George is *directly* responsible, this diagnosis also has a strange consequence. George, in deciding to be impulsive, would have freely given up direct responsibility for his behavior. Combining reasons-responsiveness with a judgment of indirect responsibility seems to allow this. It's not obvious that one could do this on other accounts of responsibility, at least not in the same way. With Frankfurt's version of compatibilism (1971), for example, having some second-order desires about which first-order desire one wants to be effective might be unavoidable, and thus not subject to the kind of control exhibited in the George case. For those libertarians who think we make free choices all the time, we could only accomplish this by refraining from action altogether. Even for those who think we only seldom make free choices, it's not as if we can simply ignore the ability and still make a decision anyway. But if this kind of

[21] More on the direct/derivative responsibility distinction can be found in Ginet (2000) and Kane (1996, p.39).

reasons-responsiveness really is the right control condition for responsibility, and George can be said to be only derivatively responsible, then he has freely relinquished direct responsibility for his behavior merely by ignoring reasons to act otherwise.

A different response to the George case may be that he is not morally responsible at all. In order to motivate this sort of response one might appeal to epistemic tracing conditions. In cases in which an agent fails to satisfy the conditions for responsibility at the time of action it often seems appropriate to trace responsibility back to a point at which the agent did satisfy the appropriate conditions. For example, we might think that responsibility for boorish behavior while drunk can often be traced back to a free decision to drink. It's this type of tracing that was assumed to be appropriate in the previous diagnosis—that George's responsibility for his theft can be traced back to his previous free decision to act impulsively throughout the year. Manuel Vargas, however, has developed cases to show that accounts of tracing may be more difficult to come by than previously thought (Vargas, 2005). He outlines a general knowledge condition on responsibility that many of our intuitions on tracing seem to follow:

> "(KC) For an agent to be responsible for some outcome (whether an action or consequence) the outcome must be reasonably foreseeable for that agent at some suitable prior time." (Vargas, 2005, p.274)

Vargas goes on to show that there are certain cases where, although we would like to hold agents responsible for their behavior, it would be inappropriate given this condition. The question, then, is that of whether the case of George is one of these problem cases.

It might be said that George, at the time of his resolution, could not reasonably have foreseen that he would steal from the restaurant's employees. This is certainly a

plausible claim since we couldn't expect George to know what his immediate desires will lead him to do, especially six months down the road. It could be argued, then, that he does not meet the knowledge condition since his action was not reasonably foreseeable. Of course I, and many others, will have a strong intuition that George *is* responsible for his action, if not directly then certainly indirectly. Even if tracing poses challenges in certain cases, and we ought to have a more refined understanding of tracing conditions, I suspect that any adequate account will result in George being responsible. First, we could compare George to a paradigm case of derivative responsibility—drunkenness. As mentioned above, one who freely chooses to get drunk, knowing that there is a reasonable chance he might act impermissibly, will typically still be held responsible for those drunken acts. A knowledge condition suitable for tracing should not require that the *particular* outcome is foreseeable at a prior time. Rather, it would have to require that some *range* or *type* of outcome is reasonably foreseeable. The drunk, for example, may foresee that he might punch someone, throw a beer bottle, or yell angrily at the bartender. He need not know which of these will occur, but so long as he foresees that some mild debauchery could take place then he's responsible for what comes.[22] George, we can suppose, knows that he sometimes desires to act wrongly, has no reason to think this will be different in the future, and still chooses to become impulsive. Although he couldn't reasonably expect to *steal money on the first day of May*, he could reasonably expect performing a number of relatively mildly wrong actions. His history of sometimes desiring to do bad things would give him reason to

[22] How we define the range of foreseeability will also be important. The drunk may think that a bar fight is possible, and so "violence" is foreseeable, though stabbing another patron to death may be violent but not foreseeable.

believe that this will be the case. Fischer himself, along with Neal Tognazzini, has defended a similar line of argument in response to one of Vargas' cases (Fischer and Tognazzini 2009: pp.537-39). Their contention is that requiring foreseeability in a very fine-grained way can lead to incorrect judgments of responsibility, and this seems to be what would be at issue with the George case.

Even considering foreseeability of the act, though, it seems that George could still satisfy a more fine-grained knowledge condition. We could suppose that George reasonably expected that he would perform that particular theft, perhaps even on that particular day. He may have been to that restaurant in the past (before the resolution), had the desire to take some money out of the jar, but refrained because of the security camera. It could be that he goes there every May 1st, to celebrate summer's near arrival, and so expected to have the desire to go on that day. Before making his resolution George may have thought about all of the bad desires he's had, including stealing from that jar. Since he eats there on occasion, he could have even thought to himself that once he starts acting on his immediate desires he will likely steal some money from that jar on his next trip in there. Moreover, he could have had the expectation to steal that money before walking into the restaurant that day. When the desire for food hit him on the first of May it may have occurred to him that the jar would be there and that, upon seeing it, he would likely have a desire to take some money. This wouldn't require him to view the money's being in the jar as supplying a reason for or against going to that restaurant for dinner. He could simply *expect* that when he sees the jar he will desire, and take, some money from it. Although Vargas has shown that there are problematic cases for tracing, then, George will not be among them. Any adequate characterization of the

knowledge condition, and tracing, will permit us to say that George may at least be derivatively responsible. If he is not responsible at all, it will have to be for some other reason.

Another argument for the claim that George is not responsible at all may be that a third-party observer is unlikely to discern a sane, understandable pattern of reasons-responsiveness from his behavior. An agent merely acting on impulse, it could be said, would behave so erratically that this component of the Fischer-Ravizza view would not be satisfied. Two responses to this line of argument indicate that this is not a promising route. First, it's easy to imagine that George behaves in such a way that he *does* appear appropriately reasons-responsive from a third-person perspective. He may just appear to be *a person acting on his first desires without regard for anything else*. This would be an understandable pattern of responsiveness and wouldn't necessarily appear insane or erratic. Moreover, we might suppose that George's desires are consistent enough that he appears to be acting on an appropriately coherent pattern of reasons. Even if he is not recognizing reasons in the appropriate way, and not evaluating them at all, he may be indistinguishable from someone who is.

Since his pattern of reasons won't preclude responsibility, and any challenges from epistemic conditions can be met, it seems that the view that George is not responsible at all for his theft is implausible. The view that George is merely derivatively responsible is problematic as well—it has the consequence that agents could freely relinquish direct responsibility for their actions and it's inconsistent with the diagnosis that a reasons-responsiveness theory ought to yield. A modification to the theory may resolve the latter issue but would not be easy to come by. If the modification were

introduced only to deal with cases like George, with no independent justification, it would be *ad hoc*. The modification would also need to be constructed in such a way that it explains the nature of George's responsibility but doesn't lead to incorrect assessments of responsibility in other cases. Finally, I maintain that the most intuitive view is that George is *directly* responsible for his theft, and so a modified reasons-responsiveness theory that diagnoses him as derivatively responsible would be unappealing.

1.6 Abandoning Reasons-Responsiveness

I contend that the most plausible solution to this problem is that George *is* directly responsible for his theft and that reasons-responsiveness is not a necessary or explanatory condition for responsibility. Instead, what gives him the appropriate control for blameworthiness is his ability to *settle* on performing a morally wrong action. I will say a little bit more about what it may mean to *settle* on an action, but a pressing question at this point is why reasons-responsiveness theories like Fischer and Ravizza's correctly diagnose such a large number of cases. It could perhaps be that reasons-responsiveness, rather than being a necessary and explanatory condition for responsibility, is something like an imperfect stand-in for this settling condition—that is, something of a simulacrum. Reasons-responsiveness itself may not explain responsibility but instead provides indirect evidence for a settler. The fact that an agent is receptive and reactive to reasons for doing otherwise could appear responsibility-relevant only because it makes it *seem* as though agents are settling which actions they perform and, perhaps, making a difference as to what occurs. If we were to give a rational agent the ability to

settle on actions we would expect that agent to be reasons-responsive in most cases, regardless of whether the latter was a responsibility-relevant control feature. The vast majority of the time we do deliberate about how to act, and respond to reasons for performing one action or another, and this simply goes along with the presumption that we ultimately settle on a particular course of action. The explanation in the case of George would be that this is an instance where reasons-responsiveness and this settling condition come apart. George's responsibility is preserved nonetheless, I suggest, because the control condition for responsibility is really of this more primitive sort. He is intuitively directly responsible because he still has the ability to settle which actions he performs, even if he fails to consider reasons for alternative actions.

As for what settling amounts to, I think that we have an intuitive grasp on what it means for an agent to settle on an action or decision. Some have also offered accounts of, or at least gestured at what I have in mind. Pereboom holds that "an agent settles which option for action occurs just in case she determines, not necessarily causally, which action occurs, and she makes the difference as to which action occurs" (Pereboom, 2017, p.11). Chris Franklin speaks of settling alongside "self-determination" (Franklin, 2016), and he considers this to involve something akin to Michael Bratman's "directing and governing" (Bratman, 2007). Helen Steward considers settling in terms of an agent *resolving* certain question—the *whethers, whens, wheres*, and so forth related to actions (Steward, 2012, p.39). Settling, then, will be a kind of basic mental action, and in particular one that comes about with a sufficient awareness of what's being settled. This awareness aspect would rule out that animals and young children have the sort of control required for moral responsibility, though

it'll still be easier to come by than, say, being moderately reasons responsive. One could have the *control* required for moral responsibility simply by consciously settling on an action, or one action rather than another. Accordingly, more emphasis on epistemic and ownership conditions may be required when assessing whether an agent is morally responsible. In fact, I would suggest that reasons-responsiveness is better equipped to determine whether agents satisfy *epistemic* conditions for moral responsibility. The ability to recognize and evaluate reasons would help to establish that an agent is aware of the moral significance of her actions and alternatives, especially given certain content of the reasons and mode of evaluation.

While I've argued for a rejection of reasons-responsiveness, and suggested a settling condition in its place, this is not itself an argument for incompatibilism or against compatibilism. First, the requirement that an *agent* have the ability to settle will be problematic for event-causal accounts of free will and responsibility, regardless of whether the account is incompatibilist or compatibilist. This is because, if anything at all settles which actions are performed on event-causal accounts, it will not be an agent but at most an agent-involving event.[23] Next, depending on its formulation, it may be possible for an agent-causal compatibilist account to meet a settling condition.[24] Two questions will be especially relevant here—that of whether settling necessarily involves *making a difference* and that of whether difference-making is compatible with determinism. If settling on an action does not require that one make a difference as to whether that decision occurred then it's hard to see what would prevent a compatibilist

[23] More on this can be found in Pereboom's "disappearing agent" objection to event-causal libertarianism (2014, pp.31-9, 2017).
[24] Examples of agent-causal compatibilist accounts include Markosian (1999) and Nelkin (2011).

theory from meeting this condition. On the other hand, if settling requires making a difference then causal determination may prevent anyone from meeting this condition.[25] Making a difference may require being the ultimate source of one's actions or access to genuine metaphysical alternatives, neither of which a compatibilist theory can accommodate. Carolina Sartorio, however, argues that difference-making can be secured even if determinism is true (Sartorio, 2013). Invoking the notion of absence causation, she claims that "causes make a difference to their effects, not in the sense that the effects would not have occurred in their absence, but in the sense that their effects would not have been caused by their absences" (Sartorio, 2013, p.193). If this is plausible then a compatibilist account could accommodate a settling condition that also involves difference making, resisting the incompatibilist alternative. As I've argued though, such a compatibilist account ought not include reasons-responsiveness as a part of the control condition.

1.7 Conclusion

As the different "Gene" cases show, George retaining the capacity for recognizing and evaluating reasons does not explain why he is responsible for his theft. George has the capacity, Gene and ^{F}Gene do not, and yet there can be no principled difference in our responsibility assessments between the cases. Fischer and Ravizza's view also can't account for why he *wouldn't* be considered directly responsible given that he meets all of their conditions and alternative assessments of responsibility are implausible.

[25] Steward is a notable proponent of this view (2012).

Reasons-responsiveness theories provide the wrong explanation for why George would be directly responsible and lack the resources to justify an indirect or non-responsible judgment. The explanation that I offer is that reasons-responsiveness is a condition which is often-satisfied but not necessary for responsibility. If this is true then whether or not George, or anyone, has the control required for responsibility will not turn on considerations about reasons-responsiveness. Although this could be seen as support for other compatibilist theories, I contend that there is a more plausible, and potentially neutral explanation of the case—reasons-responsiveness is a not-quite-reliable indicator of a settling agent, and *this* is why the intuition that George is directly responsible persists. Reasons-responsiveness likely has important implications regarding an agent's rational capacities but the fundamental ability to settle on an action may do all of the work required of a control condition on responsibility.

References

Bratman, Michael. (1997). "Responsibility and Planning." *Journal of Ethics* 1 (1): 27–43.

Bratman, Michael. (2005). "Planning Agency, Autonomous Agency." In James Stacey Taylor, ed., *Personal Autonomy*. New York: Cambridge University Press: 33-57

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

Fischer, John Martin. (2004). "Responsibility and Manipulation." *The Journal of Ethics* 8 (2): 145–77.

Fischer, John Martin. (2005). "Reply: The Free Will Revolution." *Philosophical Explorations* 8 (2): 145–56.

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

Fischer, John Martin. (2009). *Our Stories: Essays on Life, Death, and Free Will*. New York: Oxford University Press.

Fischer, John Martin. (2012a). *Deep Control: Essays on Free Will and Value*. Oxford: Oxford University Press.

Fischer, John Martin. (2012b). "Semicompatibilism and Its Rivals." *Journal of Ethics* 16 (2): 117–43.

Fischer, John Martin, and Neal A. Tognazzini. (2009). "The Truth about Tracing." *Nous* 43 (3): 531–56.

Frankfurt, Harry. (1969). "Alternate Possibilities and Moral Responsibility." *The Journal of Philosophy* 66 (23): 829–39.

Frankfurt, Harry. (1971). "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68 (1): 5–20.

Franklin, Christopher Evan. (2016). "If Anyone Should Be an Agent-Causalist, Then Everyone Should Be an Agent-Causalist." *Mind* 125 (500): 1101–31.

Ginet, Carl. (2000). "The Epistemic Requirements for Moral Responsibility." *Philosophical Perspectives, 14*, 267-277.

Hunt, David P. (2005). “Moral Responsibility and Buffered Alternatives.” *Midwest Studies in Philosophy* 29 (1): 126–45.

Judisch, Neal. (2005). “Responsibility, Manipulation and Ownership.” *Philosophical Explorations* 8 (2): 115–30.

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

Markosian, Ned. (1999). “A Compatibilist Version of the Theory of Agent Causation.” *Pacific Philosophical Quarterly* 80: 257–77.

McKenna, Michael. (2001). “Review of John Martin Fischer and Mark Ravizza’s Responsibility and Control.” *The Journal of Philosophy* 98 (2): 93–100.

McKenna, Michael. (2005). “Reasons Reactivity and Incompatibilist Intuitions.” *Philosophical Explorations* 8 (2): 131–43.

McKenna, Michael. (2013). “Reasons-Responsiveness, Agents and Mechanism.” In David Shoemaker, ed., *Oxford Studies in Agency and Responsibility*, vol.1. Oxford: Oxford University Press: 151-84

Mele, Alfred R., and David Robb. (1998). “Rescuing Frankfurt-Style Cases.” *The Philosophical Review* 107 (1): 97–112.

Mele, Alfred R. (2000). “Reactive Attitudes, Reactivity, and Omissions.” *Philosophy and Phenomenological Research* 61 (2): 447–52.

Mele, Alfred R. (2006a). “Fischer and Ravizza on Moral Responsibility.” *Journal of Ethics* 10 (3): 283–94.

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

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

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

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

Pereboom, Derk. (2017). “Responsibility, Agency, and the Disappearing Agent Objection,” *Le Libre-Arbitre, approches contemporaines*, Jean-Baptiste Guillon (ed.), Paris, Collège de France, 2017, pp. 1-18.

Sartorio, Carolina. (2013). “Making a Difference in a Deterministic World.” *Philosophical Review* 122 (2): 189–214.

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

Shabo, Seth. (2005). “Fischer and Ravizza on History and Ownership.” *Philosophical Explorations* 8 (2): 103–14.

Steward, Helen. (2012). *A Metaphysics of Freedom*. Oxford: Oxford University Press.

Strawson, Peter. (1962). “Freedom and Resentment.” *Proceedings of the British Academy* 48: 187–211.

Vargas, Manuel. (2005). “The Trouble with Tracing.” *Midwest Studies in Philosophy* 29 (1): 269–91.

Watson, Gary. (1975). “Free Agency.” *Journal of Philosophy* 72 (8): 205–20.

Watson, Gary. (2001). “Reasons and Responsibility.” *Ethics* 111 (2): 374–94.

CHAPTER 2

FRANKFURT CASES AND THE PRINCIPLE OF ALTERNATIVE POSSIBILITIES: A DILEMMA FOR BUFFERED ALTERNATIVES

2.1 Background

That the ability to do otherwise is necessary for moral responsibility is at the same time one of the most intuitive and highly contested claims in the contemporary free will debate. The intuitive plausibility becomes clear when considering our normal blaming practices and what's typically countenanced as an excuse. Epistemic failures concerning the nature and moral significance of an action can exempt an agent from moral responsibility, but so too can lacking the appropriate sort of control. Explanations like "*he was forced*," "*she had no choice*," and "*there was no alternative*" are often used to excuse agents from blameworthiness. And appropriately so, as these sorts of explanations aim to indicate that the relevant action was not performed freely. The appeal, at least on the surface, seems to be to the claim that the agent could not have acted otherwise and thus is not morally responsible. In fact, on one intuitive understanding this is simply *what it means* to act freely in the most general sense—to be able to choose an action among alternatives. As a condition for moral responsibility this has come to be known as the "Principle of Alternative Possibilities" (PAP). Characterized in terms of blameworthiness the principle can be formulated as follows:[26]

[26] The motivation for formulating the PAP in terms of only blameworthiness is twofold—first, all of the cases I'll be dealing with here concern blame rather than praise. Second, some theorists maintain that the conditions for moral blame and praise are asymmetric, and in particular that while blame may require the

PAP: An agent is blameworthy for an action *X* only if she could have knowingly performed some alternative action, thereby avoiding blameworthiness for *X*

In this form the principle captures that essential intuitive idea that the ability to do otherwise is necessary for blameworthiness while also incorporating an important historical caveat—the need for *robust* alternatives. As it turns out, alternative possibilities can come cheap, especially those outside the agent's control or those with no moral significance. In order to be plausible, the PAP must require only alternatives that are robust enough to ground judgments of responsibility.[27] The requirements that an agent can "knowingly" perform an alternative and that the alternative matters to blameworthiness capture this aspect.

Despite its intuitive force and apparent connection to our ordinary blaming practice, the PAP is far from uncontroversial. The flood of challenges to this requirement began with the publication of Harry Frankfurt's "Alternative Possibilities and Moral Responsibility" (1969). It was there that Frankfurt proposed a counterexample, meant to both demonstrate that the PAP is false and explain why it appeared true in the first place. His refutation of the principal rests on the claim that agents can be in situations, commonly called "IRR" scenarios,[28] in which an action is rendered inevitable by external factors and yet still brought about by the agent herself:

> There may be circumstances that constitute sufficient conditions for a certain action to be performed by someone that therefore make it impossible for the person to do otherwise, but that do not actually impel the person to act or in any way produce his action. A person may do something in circumstances that leave

ability to do otherwise praise does not (see, e.g., Susan Wolf, 1980, Dana Nelkin, 2011). I take it to be a further question whether or not the PAP also extends to praiseworthiness and won't take up that issue here.

[27] On the motivation for "robust" alternatives see John Martin Fischer (1994), for a full criterion for robustness see Derk Pereboom (2014, pp.10-14).

[28] The notation "IRR" (an abbreviation for "irrelevant," in reference to the external factors ensuring an action in Frankfurt cases) comes from David Widerker (1995).

> him no alternative to doing it, without these circumstances actually moving him or leading him to do it—without them playing any role, indeed, in bringing it about that he does what he does. (Frankfurt, 1969, p.830)

The key insight is meant to be that the factors which make an action unavoidable need not coincide with an agent's reasons for performing that action. Factors which determine that an action occur may be wholly irrelevant to the agent as she may not even be aware of their existence. The action may flow from her own desires and motivations just as it would have had there been no external factors rendering it unavoidable. Frankfurt's own example has these features, a condensed version of which (along with the intended upshot for the PAP) is given by Michael McKenna:

> Black wants Jones to shoot Smith by a certain time. Black would much prefer that Jones do the shooting on his own. But wishing to ensure the outcome that Jones shoots Smith by the time in question, Black covertly arranges conditions that allow him to manipulate Jones into shooting Smith should Jones show any indication that he will not shoot Smith by the time in question. As it happens, Jones shoots Smith on his own by the crucial moment. Black never intervenes. PAP is refuted by a counterexample to it; Jones is morally responsible for shooting Smith, though, due to Black's arrangements, Jones cannot do other than shoot Smith. (McKenna, 2008, pp.771-772)

This appears to be a genuine IRR scenario since the conditions guaranteeing the action (Black's secret plan) play no role in bringing about the action for which Jones is morally responsible (the shooting of Smith).

The success of this case, and potentially the success of any genuine IRR scenario, would falsify the PAP. [29] Jones is intuitively blameworthy for shooting Smith and yet it's not the case that he could have knowingly performed an alternative action—

[29] While I do believe that a genuine IRR scenario would undermine the PAP, this need not be taken for granted. It could be argued that the mere possibility of a genuine IRR scenario is not itself enough to falsify the PAP (see, e.g., Widerker, 2000, pp.188-191).

external, nonproductive forces made his action inevitable. Of course, demonstrating that the PAP is false was only part of Frankfurt's goal. As I said, he also believed that by considering IRR scenarios we could gain insight into why it was thought true in the first place. Very often the conditions which render an action inevitable *are* also reasons contributing to an agent's performing that action. In those kinds of situations blameworthiness is excused, and rightfully so. However, Frankfurt holds that this is because "we understand the person who offers the excuse to mean that he did what he did *only because* he was unable to do otherwise" (Frankfurt, 1969, p.838). This initially appears to be an appeal to the PAP and without a reason to doubt it we're inclined to stop there. In light of Frankfurt's proposed counterexample, however, we're meant to see that excuses of this sort are actually an appeal to something else. That is, that the agent was not able to act in the way they wanted, in accordance with their own reasons and desires. This, according to Frankfurt, is independent of whether the agent had genuine alternatives. Returning to those excusing explanations with this in mind, "*he was forced"* is excusing not because of a lack of alternatives, but because and only insofar as the agent was forced to do something he didn't want to do. An agent can avoid blameworthiness when "*she had no choice"* only when the factors removing her choice also explain why she acted as she did. "*There was no alternative,*" in its theoretical sense at least, is no excuse at all.

If Frankfurt is right then the PAP fails as a condition for moral responsibility and there must be some other principle that helps determine whether agents have the right sort of control for blameworthiness in particular situations. This is significant, no doubt, but the stakes are even higher. In contemporary discourse on the subject, free

will *just is* whatever turns out to be the kind of control required for moral responsibility. Control conditions can be more fine-grained, but in terms of a general characterization of what it means to have the control required for moral responsibility the PAP is the natural, intuitive option. Or it was, at least. With Frankfurt's proposed counterexamples and the many other "Frankfurt-style" cases that followed, there are now two main alternatives in characterizing the control condition for moral responsibility (and thus free will). "Leeway theorists" still endorse the PAP, maintaining that access to alternative possibilities is required for moral responsibility. They can no longer lean too hard on the intuitive appeal of the principle and, of course, acquire the burden of dispelling any Frankfurt-style cases that may challenge it. On the other side are "source theorists." Considerations of Frankfurt-style cases lead many to think that, rather than access to alternative possibilities, what's really required for moral responsibility is being *the appropriate source of one's actions*. These source theorists can be compatibilists about free will and determinism or incompatibilists, though there will be disagreement over what amounts to "appropriate" sourcehood. On the incompatibilist side, skeptics and libertarians think that moral responsibility requires being the *ultimate* source of one's actions, undetermined by forces beyond one's control. Compatibilists have a more modest, and arguably more attainable, understanding of appropriate sourcehood.[30] Being the appropriate source of one's actions will involve things like being free from external constraints, having the ability to act in accordance with reasons, or having elements of one's psychology harmonize in the right sort of way. The introduction of

[30] Though it may be easier, compatibilists don't have to opt for a modest understanding of appropriate sourcehood. McKenna, for example, has argued that ultimacy could be understood in a way that compatibilist theories can accommodate (2008b, esp. pp.198-202).

Frankfurt-style cases, then, frames the discussion of free will and moral responsibility differently, and engenders a new potential understanding of what it means to have control. Although skeptics, libertarians, and compatibilists alike can adopt this new understanding, Frankfurt-style cases are not without significant implications for the debate *between* compatibilists and incompatibilists.

Frankfurt-style cases, if successful, also undermine what have historically been important arguments motivating incompatibilism about determinism and free will. It's very easy to see why determinism might threaten free will with the PAP in the background—if all of our decisions and actions are the inevitable result of causal processes stretching indefinitely into the past then there's no metaphysical sense in which we can choose and act differently. Peter van Inwagen famously captured this in his "Consequence Argument," arguing that, because we can't change the past or laws of nature, we can't render propositions about our actual decisions and actions false and thus cannot do otherwise than we actually do (1975).[31] With the status of the PAP up in the air it's unclear how far this conclusion gets an incompatibilist. It *may* mean that determinism rules out free will and moral responsibility, but that hangs on whether Frankfurt-style cases work as intended. This motivates the need for more direct arguments moving from the truth of determinism to a lack of free will and moral responsibility (e.g., van Inwagen 1983). It also increases the importance of other, more elaborate incompatibilist arguments like "manipulation" arguments. The strategy of a manipulation argument is to claim that manipulated agents are not free or responsible,

[31] Similar incompatibilist arguments appealing to the PAP, more or less directly, include Ginet (1966) and Wiggins (1973).

there is no relevant difference between a manipulated agent and a determined agent, and thus a determined agent is not free and responsible.[32] While arguments like these don't directly rely on alternative possibilities, there's still a sense in which the failure of the PAP could make them more difficult to assess. Manipulation arguments crucially rely on our intuitive assessments of moral responsibility, and in particular that we think manipulated agents don't deserve blame for actions related to their manipulation. If this intuitive judgment is the result of a commitment to the PAP, consciously or unconsciously, then a key premise in manipulation arguments could be undermined.[33] More generally, the failure of the PAP might also make the success of manipulation arguments, and indeed all arguments against compatibilism, even more vital for the incompatibilist. Without the intuitive force of the need for alternative possibilities, and with many plausible source compatibilist alternatives, it seems that incompatibilists now bear more of the dialectical burden than was shared in the past.

Fortunately for incompatibilists, no Frankfurt-style case proposed yet has been successful in establishing a genuine IRR scenario. Or so I'll argue. Traditional Frankfurt-style cases (including Frankfurt's own and others like it) fall to what's known as the "dilemma defense." Even those Frankfurt-style cases that are constructed to avoid the dilemma defense have plausible responses that at least make it unclear whether or not they succeed. One type of case though, which I'll call "buffered alternatives" cases,

[32] Examples of manipulation arguments can be found in Pereboom (2001, 2014) and Mele (1995).

[33] This concern benefits the less-common "hard-line" response to manipulation arguments (see, e.g., McKenna, 2008c, 2014), which admits that there is no relevant difference between ordinary determined agents and certain manipulated agents but holds that even manipulated agents are free and responsible (at least when they satisfy the relevant compatibilist conditions). The alternative, "soft-line" response maintains that there *is* a relevant difference between manipulated agents and ordinary determined agents.

remains without a well-established response from leeway theorists. My main purpose here will be to argue that buffered alternatives cases are also unsuccessful, as they too force the Frankfurt defender into a dilemma. In section 2 I'll describe both how traditional Frankfurt-style cases fail in light of the original dilemma defense and the issues facing certain subsequent kinds of cases. In section 3 I'll present a buffered alternatives case and argue that it falls to a dilemma, though of a different sort than what challenges traditional Frankfurt-style cases. Section 4 will have my responses to potential objections to the dilemma and in section 5 I'll outline what I take to be some implications for the larger debate, including source vs. leeway theory.

2.2 Traditional Frankfurt-Style Cases and Beyond

When Frankfurt laid out his initial argument against the PAP he established a formula for counterexamples—take an agent who performs a bad action of her own accord and add in an ensuring condition, typically a counterfactual intervener, which would have forced the action had the agent failed to act on her own. In order for a Frankfurt-style case to be successful it must be established both that the agent is blameworthy for the relevant action and that the agent couldn't have done otherwise with respect to that very action. This general structure is fairly easy to come by, even in such a way that these two conditions for success appear satisfied. The difficulty, however, comes in the details. Clarifying the nature of the intervention, determining what *kind* of action should be assessed, and assuring a dialectically neutral starting point all make constructing a successful Frankfurt-style case far more difficult.

One requirement of Frankfurt-style cases is that the agent involved has libertarian

free will of whatever sort the incompatibilist prefers.[34] This is because it can't be expected that the incompatibilist judge an agent to be morally responsible under determinism, regardless of the distinction between irrelevant external forces and the internal desires and motivations of the agent. If determinism is true, on an incompatibilist view, the agent will never be blameworthy for her actions since even her own internal psychology is just as much a product of causal forces beyond her control as anything external. This is important since incompatibilists, for the most part, are the main target of Frankfurt-style cases. While there were, pre-Frankfurt, many classical compatibilists that made use of an ability to do otherwise (e.g., Hume, Moore, and Ayer), very few modern compatibilists are leeway theorists.[35] In order to get the argument off the ground dialectically, then, it has to be assumed that the agent in a Frankfurt-style case acts undetermined, with libertarian free will.[36] Another feature that's important for Frankfurt-style cases is that the action for which the agent is blameworthy must be of a fundamental sort. While we may usually associate 'actions' with behaviors, or the carrying out of intentions, when it comes to moral responsibility the most fundamental kinds of actions are simple *decisions* or *intention-formations*.

[34] For the libertarian this will involve the agent-casual, event-causal, or non-causal power to bring about actions in a way that's free from deterministic processes. For the skeptic it will involve whichever of these libertarian theories *would be* enough for moral responsibility, or perhaps whichever is taken to be most plausible.

[35] Exceptions include, at least for blameworthiness, Nelkin (2011) and Wolf (1980), but also Vihvelin (2004). Each of these accounts, however, understand the ability to do otherwise differently than the typical incompatibilist. While leeway incompatibilists take the ability to do otherwise to be access to genuine metaphysical alternative possibilities, leeway compatibilists take it to be a practical ability of a certain sort.

[36] There is *something* strange about the dialectical claim that libertarian agency ought to be assumed to avoid the Frankfurt defender begging the question. After all, the reason that the leeway incompatibilist believes that determinism threatens free will and responsibility is rooted in the PAP, the very principle that's being contested. I'll set this concern aside here, though for an explanation of the issue, and an argument that the incompatibilist *is not* entitled to this assumption, see Haji & McKenna (2004).

These most basic actions are what ground blameworthiness, while responsibility for executing actions would merely be derivative from the decision.[37] Thus in Frankfurt-style cases the action being morally evaluated ought to be the decision itself. Finally, the mechanism by which counterfactual intervention would be triggered needs to be made clear in these cases. In order to assess whether an agent has alternative possibilities relative to a particular decision it has to be known how and when the counterfactual intervention would take place. Traditional Frankfurt-style cases rely on a "prior sign" indicating that the agent will not decide to act on her own. This triggers intervention, typically involving some sort of covert manipulation that forces the desired action only when the sign appears. The prior sign, combined with the assumption of libertarian free will and responsibility for basic decisions, is what opens traditional Frankfurt-style cases to the dilemma defense.

The dilemma defense, articulated independently by Kane (1985), Widerker (1995), and Ginet (1996), focuses on the prior sign present in traditional Frankfurt-style cases. In our version of Frankfurt's own case the prior sign is left implicit, but it's still very much a key feature of the scenario. Black will intervene if Jones shows "any indication that he will not shoot Smith by the time in question" and so there is *something* that would reveal to Black that he must force Jones to shoot Smith. For simplicity we can take the prior sign to be a single, visible event—say, Jones' eye twitching.[38] Adding this

[37] One motivation for this claim is that it removes the potential for a certain kind of moral luck. Two agents may make the same decision with the same intentions yet differ in how or whether the decision is carried out. In cases like these, properly clarified, it seems most plausible to say that they are equally morally responsible *because* their most basic actions were the same.

[38] Other traditional Frankfurt-style cases (e.g., Fischer, 1994) use an explicit prior sign for triggering intervention.

into the case we can say that Black will stand by, prepared to intervene only if he sees Jones twitch by a specific time. Jones doesn't twitch and instead decides to shoot Smith of his own (libertarian) free will.[39] He's blameworthy, then, for the *decision* to act as he does since he acted of his own accord while Black did nothing. The starting point for the dilemma that arises can be framed as a question—how reliable is the twitch as a prior sign for intervention?

On the one hand, if the twitch is a perfectly reliable indicator that Jones will not decide to shoot Smith then this spells trouble for the claim that Jones freely decides on his own in its absence. In order to guarantee that Black needs to intervene, the twitch must be sufficiently linked to the decision. The details of how the twitch connects to the later decision can be outlined in a number of ways, but if this sign is truly reliable then the two must be deterministically linked. It can't be, for example, that the twitch only makes it uncertain what Jones will do, or even that it makes it very likely he won't decide to act on his own. A reliable sign needs to be just that—reliable. The reason this is a problem for traditional Frankfurt-style cases is that it doesn't seem that a reliable prior sign is compatible with a later, free decision. After the time at which Jones fails to twitch his decision to shoot Smith is already determined in virtue of whatever makes the twitch a reliable sign. It's simply inconsistent with libertarian free will that a decision can be both guaranteed by events in the past and an agent's own. If this is how the case is filled in, then, incompatibilists can't be expected to have the intuition that Jones is blameworthy for acting as he does. Since a Frankfurt-style case needs an agent

[39] The nature of this decision will again depend on one's preferred account of libertarian agency. We may say, for example, that Jones agent-causes himself to decide as he does, or that he was the subject of an indeterministic agent-involving event.

who is intuitively blameworthy to succeed, the "perfectly reliable" route won't result in a counterexample and the PAP survives. If the sign is not perfectly reliable, on the other hand, then traditional Frankfurt-style cases run into a different problem. A twitch that merely makes it *very likely* that Jones will not decide to act on his own may be pretty effective for Black's purposes but it's not enough to rule out alternative possibilities. Given libertarian free will, Jones will still have the power to act on his own or refrain from acting at the time of the decision regardless of a sign that reveals his inclinations.[40] Unlike with a reliable sign he'll be blameworthy for shooting Smith, but only because he was the undetermined initiator of his action. So on either route, a reliable sign or an unreliable sign, traditional Frankfurt-style cases fail to satisfy one of the two criteria for success, intuitive blameworthiness or the removal of alternative possibilities.

Subsequent Frankfurt-style cases have attempted to evade the dilemma defense by constructing scenarios where the sign for intervention can't be exploited in this way, or scenarios without a prior sign altogether. Eleanor Stump, for example, developed a case where potential intervention is queued *during* the temporal duration when the indeterministic decision takes place (1996). Supposing that decisions can involve multiple neural events that are both discrete and reliable, the potential intervener can look for certain signature neural firings that constitute a particular choice and know when to intervene. Since this sign would be a part of the decision itself it's meant to avoid the issues that arise from a *prior* sign. Al Mele and David Robb remove the sign

[40] Most libertarian accounts have seen the need to make room for psychology that makes it more likely that an agent will perform one decision rather than another. Chisholm, for example, adopted the Leibnitzian view that desires "incline without necessitating" (1964). Kane, an event-causal libertarian, allows for this in a more derivative way, as agents create their wills through "self-forming actions," paradigmatically occurring through torn decisions (1996).

altogether, instead stipulating a covert, deterministic process that culminates in forcing the desired decision only if their agent, Bob, fails to decide on his own by a particular time (1998). This process isn't linked to Bob's own indeterministic decision in the way that a reliable sign is in traditional Frankfurt-style cases and so shouldn't threaten the ability to act in an incompatibilist-friendly way. When Bob decides to act on his own, then, he's clearly blameworthy for having done so. Moreover, unlike with an unreliable prior sign, the deterministic process isn't fallible in a way that leaves open robust alternative possibilities. Strategies like these adjust the style of intervention in such a way that the dilemma defense is harder, or even impossible to press. Still, the challenges they face leave it unclear whether they represent genuine IRR scenarios. Stewart Goetz argues that Stump's case still fails to avoid the dilemma defense, or at least a similar version of it (1999). While her sign is internal, and even part of the indeterministic decision, it's still either deterministically linked to the completion of the decision or not. Arguably, then, the same challenge arises. Mele and Robb avoid the dilemma defense but ultimately rely on a controversial metaphysical principle. There are multiple ways Bob's action can come about depending on whether and when he decides on his own to act or whether and when the deterministic process causes him to act. The possibility that potentially represents an IRR scenario will be if Bob decides on his own to act at *exactly* that time when the deterministic process is set to cause him to decide. In this scenario Bob's decision is said to *preempt* the deterministic process, rendering it inert while his indeterministic decision plays the only causal role. This feature is crucial for success—if the decision were overdetermined then it would be unclear whether Bob is blameworthy, yet if the deterministic process caused the decision then Bob simply

wouldn't be blameworthy. The concern, however, is that this sort of "occurrent," or "trumping" preemption may not be metaphysically possible. In order to eliminate alternative possibilities it must also be the true that, if Bob actively decides *not* to act at the specified time, then the deterministic process preempts *that* decision instead. A stipulation that preemption reverses based the content of Bob's decision is certainly cryptic and success of the case hangs on whether it's conceivable that a metaphysical law could accommodate and govern these two scenarios in just this way. It would be hasty to claim that it's outright impossible, but it's the Frankfurt-defender who holds the burden of establishing that it's coherent.[41]

So traditional Frankfurt-style cases fail in light of the dilemma defense and other kinds of cases, if they avoid the dilemma at all, face notable challenges. It's safe to say that the PAP stands strong against these former cases and at least survives, for the time being, against versions like Stump's internal sign case and Mele and Robb's no-sign case. This is good for the leeway theorist but of course it's not the most encouraging progression. It only gets more difficult from here as well, since a newer kind of Frankfurt-style case, developed by Pereboom (2001) and then Hunt (2005), is even harder to challenge. These cases, "buffered alternatives" cases, introduce a feature that allows them to side-step the issues that arise from a temporally specific prior sign or a deterministic form of intervention.

[41] Mele and Robb attempt to address this concern in a follow-up to their case (2003). Jonathon Schaffer also utilizes trumping preemption in the more general context of counterfactual theories of causation (2000).

2.3 A New Dilemma for Buffered Alternatives

What makes buffered alternatives cases so inventive is that they're able to keep the prior sign that queues intervention, but they're set up in such a way that the sign is not deterministically linked to the ultimate decision. Because of this the sign can be reliable but still avoid undermining the agent's libertarian free will. The feature that allows for this is the introduction of a necessary-but-not-sufficient condition that the agent must satisfy in order to choose an alternative course of action. Instead of looking for an involuntary sign or setting up a deterministic process, the counterfactual intervener simply looks to see if and when the agent meets the necessary condition. There's no specific time for intervention and neither the counterfactual intervention nor the agent's decision is determined. It's completely up to the agent whether or not he satisfies the necessary condition and he's free to do so at any point up until the relevant decision. The dilemma defense devised by Kane, Widerker, and Ginet can't be posed because *everything*—the counterfactual intervention, the potential satisfaction of the necessary condition, or the ultimate decision—is indeterministic and, directly or indirectly, brought about by the agent.

In Pereboom's buffered alternatives case we find an agent named "Joe" who will ultimately cheat on his taxes. The most recent version is titled "Tax Evasion 2":[42]

> Joe is considering claiming a tax deduction for the registration fee that he paid when he bought a house. He knows that claiming this deduction is illegal, but that he probably won't be caught, and that if he were, he could convincingly plead ignorance. Suppose he has a strong but not always overriding desire to advance his self-interest regardless of its cost to others and even if it involves illegal activity. In addition, the only way that in this situation he could fail to choose to evade taxes is for moral reasons, of which he is aware. He could not,

[42] The initial version of "Tax Evasion" is presented and defended in Pereboom's *Living Without Free Will* (2001, pp. 18-33).

> for example, fail to make this choice for no reason or simply on a whim. Moreover, it is causally necessary for his failing to choose to evade taxes in this situation that he attain a certain level of attentiveness to moral reasons. Joe can secure this level of attentiveness voluntarily. However, his attaining this level of attentiveness is not causally sufficient for his failing to choose to evade taxes. If he were to attain this level of attentiveness, he could, exercising his libertarian free will, either choose to evade taxes or refrain from so choosing (without the intervener's device in place). However, to ensure that he will choose to evade taxes, a neuroscientist has, unbeknownst to Joe, implanted a device in his brain, which, were it to sense the requisite level of attentiveness, would electronically stimulate the right neural centers so as to inevitably result in his making this choice. As it happens, Joe does not attain this level of attentiveness to his moral reasons, and he chooses to evade taxes on his own, while the device remains idle. (Pereboom, 2014, p.15)

Something to note with this case is that, even though Joe has an alternative possibility in that he can consider the moral reasons against claiming the deduction, Pereboom argues that this is not a *robust* alternative possibility. A robust alternative must be "relevant per se to explaining why an agent is morally responsible for an action" (Pereboom, 2014, p.10). Joe, with no knowledge of the intervener, is not sensitive to the fact that had he considered the moral reasons he would not be blameworthy. As far as he is concerned, had he reached the requisite level of moral attentiveness he could still have freely chosen either to evade taxes or not. The presence of the neuroscientist, then, removes the kind of alternative possibilities that are at issue with the PAP. The necessary-but-not-sufficient condition that acts as a buffer between Joe's decision to cheat on his taxes and any robust alternative is "reaching a level of moral attentiveness," which is fully under his own control.[43] The necessity is meant to secure the elimination of alternative possibilities while the lack of sufficiency ensures that he maintains

[43] While it's under his control in the sense that he can, at any time, voluntarily reach the requisite level of attentiveness, this doesn't rule out that he reach this level *involuntarily*. This feature of the case won't impact my discussion and thus I'll speak of meeting this condition in terms of a voluntary act.

libertarian free will throughout the process. After Joe fails to adequately consider the moral reasons and ultimately cheats on his taxes, he'll be blameworthy for having done so. Since a successful Frankfurt-style case requires only a blameworthy agent who doesn't have alternative possibilities, this appears to be a plausible counterexample to the PAP. While the traditional dilemma defense can't be posed, however, a different kind of dilemma arises given the nature of buffered decisions.[44]

In my view Joe will likely be blameworthy when he cheats on his taxes as there's no reason to think that his action isn't ultimately the result of libertarian agency. The dilemma I'll pose, then, concerns only whether *and when* he may have had alternative possibilities relating to the decision for which he's blameworthy. Once again, a useful starting point for introducing the dilemma is a question—what explains why "the only way that in this situation [Joe] could fail to choose to evade taxes" is by considering moral reasons against it? That is, what is it about him that makes his default state one in which he will evade taxes rather than one in which he will not? After all, we can easily imagine a parallel agent, Elaine, who will choose *not* to take the deduction unless she voluntarily decides to consider *selfish* reasons to evade taxes. At that point she could, absent a counterfactual intervener, freely decide whether or not to take the deduction. I don't think it would be controversial to suggest that *something* in their respective histories or psychologies must explain the difference between Joe and Elaine's contrasting dispositions. In fact, we get a hint of this in the Tax Evasion case itself, as it's noted that Joe has a strong yet defeasible desire to advance his self-interest,

[44] While I take Pereboom's "Tax Evasion" as a model, the dilemma I develop applies to all versions of buffered alternatives cases, including, for example, Hunt's "Revenge" (2005).

without consideration for the law or others. With this characterization in mind, I turn to the first horn of the dilemma.

When Joe sits down with his tax form and comes to the section regarding deductions, assuming that he hasn't reached the requisite level of moral attentiveness, there are two ways that events can proceed. First, he may not have previously thought about how he would fill out his form and, upon seeing an opportunity to claim his home registration fee as a deduction, does so without putting any thought into the matter. He's aware that he could consider moral reasons against it, of course, but he simply does what comes instinctively to him and cheats on his taxes. Given an agent who, by nature, is inclined towards advancing his self-interest, this seems to be a very plausible elucidation of the case. Actions like these are even familiar in non-moral contexts. A habitually inattentive driver may take a wrong turn after failing to pay attention to the GPS even though he's aware that the directions are just a glance away. We shouldn't, at least upon reflection, fault him for taking the wrong turn itself. Instead, it would be more appropriate to admonish him for being habitually inattentive. On this understanding of Tax Evasion, the same would apply to Joe. He is intuitively blameworthy, though not for the action of claiming the deduction *per se*. Rather, the action merely flows from his character in such a way that, if he is in fact blameworthy, it's only insofar as he's responsible for his character. Pereboom establishes that Joe has no robust alternative possibilities relative to the act of cheating on his taxes but *not* that he had no alternative possibilities when forming his character. Filled in in this way, then, Tax Evasion doesn't pose a challenge for the PAP.

The only other way that events can proceed represent the second horn of the

dilemma. When Joe approaches his tax form he may have, at some earlier time, consciously formed the intention to cheat on his taxes. This wouldn't guarantee that he would ultimately take the deduction since he could, at any point, consider the moral reasons against it. Still, Joe's will could be *provisionally* set in this way. So we might imagine that Joe, when he awoke that morning, provisionally decided to claim the illegal deduction. Then, when he reaches that section of his tax form, he carries out his prior intention to claim the home registration fee. This too is a familiar form of action in various real-world scenarios—decisions to eat at a certain restaurant, meet with friends, or watch the sunset can all be made prior to the actual action and, importantly, without reconsideration of whether to follow through on those plans. Given the stipulations of Joe's case we can assume that he never reconsidered his decision to take the deduction. In order to consider "changing his mind," as it were, he would have first had to adequately consider the moral reasons, which he fails to do. In this scenario, as in the last, Joe won't be responsible merely for taking the deduction at the time he does so. Rather, if he's blameworthy for evading taxes, it's in virtue of his having carried out his prior intention. The fact that Joe could have freely chosen to consider the moral reasons explains why his action wasn't causally determined but doesn't establish *which* component of the course of action he's fundamentally blameworthy for. The most basic action relating to his transgression is the provisional decision and not his carrying out of that prior intention. Like on the previous horn, it hasn't been established that Joe had no alternative possibilities relative to the decision for which he's blameworthy. On this understanding of Tax Evasion the PAP survives once again.

Of course, there are two important claims I've thus far left undefended, both of

which require justification. First, I claimed that, given the details of Tax Evasion, the two scenarios I outline are the only two ways that the case can proceed. One natural alternative that may come to mind is that Joe somehow decides, in a fundamental way, to cheat on his taxes at the time at which he does so. This would be more in line with previous Frankfurt-style cases and it's the sense one gets given the way Pereboom describes the conclusion of the case. Now, if this decision were simply a conscious affirmation to take the deduction with no deliberative content then it's hard to see how this isn't just an expression of his character. There's nothing that distinguishes this type of decision from, say, when the pathological liar is presented with a new opportunity to lie. The intention to evade taxes or to lie may be formed shortly before the corresponding action takes place but it's not the type of intention that one is directly morally responsible for. A similar thing can be said if an alternative flashes through Joe's mind *in a particular way*. We might imagine, for example, that as Joe comes to the deduction section of his tax form it occurs to him that some people might not take the deduction. He may even be consciously aware that it would be possible for *him* not to take the deduction. If none of these thoughts provoke any consideration of not taking the deduction then his decision to evade taxes is again simply an expression of his character.[45] On the other hand, given that Joe hasn't become adequately attentive to moral reasons it's not possible for him to give real, deliberative consideration to an alternative. He could potentially consider *how* to go about cheating on his taxes, but not *whether he will* cheat on his taxes. Even if he could deliberate, the moment he gives

[45] Or, alternatively, the carrying out of a prior, provisional decision. As I would argue, of course, this too doesn't help avoid the dilemma.

genuine consideration to an alternative it no longer makes sense to say that he couldn't have done otherwise. If any deliberative weight is given to an action other than evading taxes then this itself will be a robust alternative. The safeguard against alternative possibilities is the necessary condition and there's nothing else in the case to guarantee that Joe evades taxes. Thus, without deliberation the case falls back into the dilemma and with deliberation, if the case were modified to allow it, alternative possibilities remain. Another route that avoids filling in the case in the ways I've outlined would be to simply reject the call for further explanation. I argued that there must be *something* that explains Joe's inclinations and how they differ from someone like Elaine's, but it could be proposed that it's just a brute fact that this is how Joe will decide. The problem with this response is akin to the problem that faces Mele and Robb's Frankfurt-style case—postulating a mysterious law allows for skepticism about its conceivability. In this case it's even harder for the Frankfurt defender to press because of what kinds of things are being governed by the law. Most people are unlikely to have a firm grasp on which causal laws could govern the neural underpinnings of decisions and so it's difficult to say whether trumping preemption is possible in those cases. When it comes to persons and their general decisions, though, it's far more intuitive that there's always historical and psychological underpinnings. One could *set these aside*, of course, but basing an argument on the possibility of their *inexistence* would be questionable. Outside of Joe's decision being unexplained in this way, or the case being modified to allow genuine deliberative alternatives, I can think of no way that Tax Evasion can be filled in that wouldn't fall under the dilemma-inducing scenarios I've described.

The second important claim I made was that we have no reason to think the

character-forming or provisional decision(s) didn't include alternative possibilities. In fact, I would make the even stronger claim that alternative possibilities for these earlier, responsibility-conferring decisions *cannot* be eliminated. The reason that there must have been alternatives at these prior times is that the only way to rule them out is to apply a Frankfurt-style case to them as well. Traditional Frankfurt-style cases and those like Stump's or Mele and Robb's won't work because of the familiar objections they face. The other possibility, then, would be to re-pose a Frankfurt-style case with the same buffered alternatives structure. In light of this, however, we could consider the same two possibilities—these *prior* decisions either flowed from Joe's character or were the carrying out of an even earlier provisional decision. These too would have been libertarian free choices that must have included alternate possibilities unless yet a further Frankfurt-style case is applied, and so on indefinitely. The result of applying these cases to earlier and earlier decisions will inevitably result in an initial decision with respect to which Joe had alternate possibilities, or else one that was determined by factors beyond his control.

What allows for this dilemma to be posed is the same feature that buffered alternatives cases rely on to avoid the traditional dilemma defense. The necessary-but-not-sufficient condition for doing otherwise blocks alternative possibilities, but simply plugging in actions with moral worth leave too much unexplained. Providing explanations like the ones I've outlined above end up shifting the locus of moral responsibility in such a way that alternative possibilities aren't ruled out for the actual responsibility-conferring decisions. What's going on in Tax Evasion if the case proceeds in either way is that Joe establishes a locus of moral responsibility at a time

prior to when he claims the illegal deduction. Given the assumption of libertarian free will and no deterministic forces in play, this isn't a once-and-for-all establishment—the act of reevaluating the type of person he wants to be or reconsidering his provisional decision would allow the locus to shift. By making Joe's psychology such that a necessary condition is required (yet unfulfilled) for either of these things, however, Pereboom has ensured that the locus *doesn't* shift. Blameworthiness will attach to the character-forming or provisional decision(s), neither of which we have reason to think didn't involve alternative possibilities.

2.4 Defending the Dilemma

Having laid out my dilemma for buffered alternatives cases I'll now address some objections that warrant discussion. One such objection would involve the concern that my dilemma rests on the view that Joe is derivatively responsibility. In itself this wouldn't be inappropriate, of course, but the objection that Joe is merely derivatively responsible has been proposed by Widerker and addressed by Pereboom. Widerker contends that Joe is blameworthy, not for the act of evading taxes directly, but for failing to be more attentive to the moral reasons (2006, pp. 173-4). The latter is an action that could've been performed, or not, without interruption by the counterfactual intervener. Since a challenge to the PAP needs to concern basic, non-derivative decisions, blameworthiness that traces to a different decision won't constitute a counterexample. Pereboom's response is twofold—first, that it seems Joe is being considered derivatively responsible because "only relative to this decision does Joe have a robust alternative possibility" (Pereboom, 2014, p.19). The charge is that the move to

derivative responsibility is motivated by the PAP intuition, which would be dialectically inappropriate. Second, Joe's situation is very unlike paradigmatic cases of derivative responsibility. It's not as if Joe knows, as one considering getting drunk might, that his actions at some earlier time will render him incapable of controlling himself at a later time. He understands throughout, it's assumed, that he could become sufficiently attentive to the moral reasons (Pereboom, 2014, p.19). I agree with Pereboom that Joe is not derivatively responsible in this way. He may *also* be responsible for failing to be more attentive to moral reasons if we think blameworthiness proliferates, attaching independently to both the relevant action and any omissions that may have prevented it. Thus, contrary to Widerker, I don't think that Joe's blameworthiness for evading taxes comes about only in virtue of having failed to consider the moral reasons against it. Still, my account of Joe's blameworthiness does suggest that his responsibility is generated by past actions. The question, then, is whether my dilemma makes use of derivate responsibility in a way that defenders of buffered alternatives cases can counter in a similar manner.

My claim, most generally, is that Joe's locus of responsibility is not at the time when he cheats on his taxes, but at some earlier time. Whether or not this constitutes "derivative responsibility" will depend on how we want to understand that term and perhaps which horn of the dilemma we're examining. On the first horn the locus of responsibility is past, character-forming decisions. Concerns about derivative responsibility could be avoided altogether by maintaining that Joe is *only* blameworthy for those prior decisions and not the instinctive act of cheating on his taxes. This would prevent doubling-down on moral judgments but it's not clear that it lines up with normal

responsibility ascriptions. It may be more intuitive to say instead that Joe *is* blameworthy for cheating on his taxes, but only in virtue of having freely formed his character in such a way that actions like these could have been expected. I'm content to concede that this is a form of derivative responsibility since it's not of the problematic sort. Aside from cases like willful drunkenness, character-forming decisions that lead to nondeliberative actions may be the most paradigmatic form of derivative responsibility. So long as an agent could have reasonably foreseen the potential effects of those kinds of decisions it's widely accepted that responsibility can trace in this way.[46] On the second horn it's less clear whether it's appropriate to label Joe's responsibility "derivative." When it comes to a provisional decision that's carried out at a later time it may make more sense to suggest that an agent is only blameworthy for that fundamental decision. We don't typically clarify that an agent is derivatively blameworthy for the physical manifestation of a decision when it occurs around the time of the more basic intention formation. Instead, we simply say that the agent is blameworthy for deciding to act as he did. What's significant for this horn of the dilemma is not the duration between the provisional decision and the act of claiming the illegal deduction, it's that Joe never reconsidered the provisional decision. Regarding decisions and the actions they produce, tracing responsibility should be a matter of finding the simple, responsibility conferring mental act. Of course, it may be that it's *always* appropriate to say that agents are responsible for their physical actions in virtue of the decisions that produced them. If this is the form of derivative responsibility that my second horn takes then it still won't be problematic. All blameworthy actions that

[46] In the literature on "tracing" see, e.g., Vargas (2005) and Fischer & Tognazzini (2009).

involve carrying out prior intentions will be derivative in this sense. Finally, on either horn, the judgment regarding the locus of responsibility in no way relies on the PAP intuition. Rather, it relies on the claim that Joe's inclination or disposition, which leads him to ultimately cheat on his taxes, could not have arisen *ex nihilo*. This concerns the general way in which actions come about but not whether alternative possibilities need to be involved.

A different concern one might have is that the necessary condition itself could provide room for deliberation, and thus reevaluation of either Joe's provisional decision or character. The counterfactual intervention only kicks in when he reaches *a certain level* of attentiveness to the moral reasons. He has control over whether he considers the moral reasons at all and presumably he could attend to them in a limited way, below the threshold for intervention. The key question will be whether "light consideration" of the moral reasons would be enough to change the locus of responsibility from where I've argued it's situated. Given the stipulation that reaching the threshold for intervention is the only way Joe could decide otherwise, it doesn't seem that light consideration could change the locus. Mere conscious awareness of the reasons against performing an action isn't itself enough to constitute deliberation. What's needed is a form of consideration that could potentially lead Joe to reevaluate, where the reasons are being attended to in a way that could give rise to reevaluation. Absent this, consideration of the moral reasons only ensures that he knows what he's about to do is wrong, or that there *exist* reasons against the action he's settled on. The problem is that if light consideration involves the ability to reevaluate then the case won't work as counterexample to the PAP. The necessary condition is what eliminates alternative possibilities because it's

only by reaching the threshold for moral attentiveness that Joe could potentially decide otherwise. Light consideration that allows for reevaluation, then, *would need to be* the threshold in order to rule out alternative possibilities. Otherwise, since Joe has libertarian free will throughout the scenario, there would be nothing to stop him from deciding not to evade taxes while lightly considering the moral reasons. In order to shift the locus of responsibility away from the provisional or character-forming decision(s) Joe must genuinely deliberate, but the point of intervention has to be prior to such deliberation.

Here we might wonder whether there's some way to preserve a locus of responsibility around the time of the action without appealing to deliberation or reevaluation. To start, it does seem necessary that plausible understandings of libertarian free agency would allow for decisions to be responsibility-conferring even if the agent's character plays a key role in the decision-making process. Randolph Clarke, for example, highlights such understandings in response to a well-known argument for responsibility skepticism (Clarke, 2005). The argument—Galen Strawson's "Basic Argument"—contends that being responsible for one's actions requires being responsible for the way one is mentally, which is impossible (Strawson, 1994). This is because when an agent acts she does so for a reason,[47] and that reason will be related to things like her beliefs and desires. If these mental states are what cause her action, then in order to be responsible for the action she would need to be responsible for those mental states. But of course, in order to be responsible for those mental states she would need to be responsible for the actions which led to them, and *those too* would have been

[47] Or, at least, when she acts in such a way that she could potentially be held directly morally responsible.

performed for a reason. As Strawson puts it, "here we are setting out on a regress" (Strawson, 1994, p.7). The ultimate causes of one's actions will trace back to genetics, environment, and maybe even randomness, but nothing for which an agent could be truly morally responsible. Clarke's response to this argument is that there are accounts of free action where, "even if it is not up to an agent how she is mentally, her action can still be up to her, she can still have a choice whether she performs that action, even when she acts for a reason (Clarke, 2005, p.16). This is certainly the case for any compatibilist account of responsibility, as these will have a different understanding of what it means to "have a choice," but Clarke has in mind libertarian accounts in particular. On "event-causal" libertarian accounts (e.g., Kane, 1996; Balaguer, 2010) the causes of actions are "agent-involving events." Although the agent may not be responsible for these events, including those related to mental states, indeterminism leaves open the possibility that the agent either act or refrain to act. On "agent-causal" accounts (e..g, Chisholm, 1964; Clarke, 1993; O'Connor, 2000) an agent-as-substance indeterministically causes actions, despite potential influence by prior events or anything else. Even if certain mental states play a role in, say, deliberation, on these accounts it's still ultimately "up to the agent" whether the relevant action comes about. On either of these kinds of accounts, the relevant actions aren't *determined* by the agent's mental states, and so there's room to stop the potential regress.[48] Now, Clarke acknowledges that whether or

[48] Although Clarke doesn't make use of the third kind of libertarianism—"non-causal libertarianism"—these accounts can attempt to side-step Strawson's challenge in a similar manner. Actions would fundamentally involve neither agent nor event causes, instead being a "causally simple mental action" with an "intrinsic actish phenomenological quality" (Ginet, 1997, p.89). In at least one sense these accounts may have a stronger response to Strawson's challenge, since how one is mentally plays *no* fundamental causal role in producing actions for which one may be held responsible.

not these kinds of libertarian accounts would stand up against a fuller defense of the Basic Argument will depend on whether they represent coherent understandings of the nature of free choice. In particular, whether the substance cause or agent-involving event that constitutes the relevant decision avoid any luck or randomness that may undermine responsibility. This concern is beyond the scope of Clarke's response, as it will be for my discussion as well. What I want to address is the implications of Clarke's discussion for my dilemma and, in the context of Frankfurt-style cases, it's appropriate to *assume* a coherent account of libertarian free agency.

The key issue at play when it comes to the Basic Argument is whether the way one is mentally necessitates particular actions.[49] Strawson assumes that it does, at least for the sorts of actions for which responsibility may apply, while Clarke suggests some accounts of free action that reject this assumption. The question, then, is what accounts like these would say about Joe's action, particularly considering the horn on which he acts as a result of his character. If he can be responsible simpliciter for the act of evading taxes, *despite* the role that his character plays in that action, then the case would be a genuine IRR scenario. There would be no regress as I suggest would occur alongside a shift of the locus of responsibility to those decisions which led to Joe's character and so the Tax Evasion would constitute a counterexample to the PAP. Of course, absent the special features of the buffered alternatives Frankfurt-style case, I think the conclusion would be clear. Joe could be responsible for the decision to evade taxes at the time at which he does so because it could still be up to him whether his desire to advance his

[49] This will be the key issue from an incompatibilist's perspective, at least. Compatibilists can accept that the way one is mentally necessitates actions while denying that this would undermines responsibility.

self-interest leads to taking the deduction. Can the same be said after the introduction of the necessary-but-not-sufficient condition? In order to address this, it's important to try to characterize—and carefully—what "up to the agent" should mean when applied to Joe. It shouldn't be, for example, that we understand it to mean "up to Joe *whether he acts or refrains from acting*" since this bakes in an alternative possibilities condition. Accordingly, I'll take "up to the agent" to mean something like, "having the ability to *settle* the decision in a way that secures responsibility".[50]

Joe's taking the deduction won't be *determined* by his character in the usual way that incompatibilists agree would threaten free choice. He is, at any point, able to choose to adequately consider the moral reasons against evading taxes and then act according to those reasons. It would also be too strong to say that his taking the deduction is *necessitated* by the combination of his inclinations and the fact that he *has yet* to adequately consider the moral reasons. That he hasn't reached the requisite level of attentiveness yet doesn't preclude his doing so later and potentially not taking the deduction. However, there is a weaker sense in which Joe's particular constitution and the necessary-but-not-sufficient condition combine to prevent its being up to him whether he evades taxes at the time he does so. That Joe is constituted such that he will take the deduction unless adequately considering the moral reasons, and that at any particular moment he has not met that condition, together make it *fixed* that he will take the deduction. We can think of this as analogous to driving a car with the cruise control feature enabled. Setting the cruise control to 55mph fixes that the car will continue at

[50] I intend this characterization to be neutral between source and leeway incompatibilist conditions for responsibility.

that speed indefinitely, regardless of any hazards or obstacles it may come across. It's not fixed once-and-for-all, of course, as one could "unfix" that condition by, say, hitting a "cancel" button or pressing the brakes. What's significant, though, is that it's *no longer* up to the driver that the car is traveling 55mph until an action is taken that unfixes that speed. The same applies to Joe as he's set to evade taxes when the opportunity arises unless he takes action to potentially reconsider, thereby unfixing that course of events. To stretch this analogy further, we can imagine Tony, who's driving a Saab convertible on the highway and has set his cruise control to 56mph. However, the speed limit is 55mph and, as Tony comes around a bend in the road, a fastidious police officer "clocks" him for a speeding violation. Certain assumptions will be useful, of course—that he knew his cruise control was set above the speed limit, that he understands that speeding is a violation, and that he could have disabled the cruise control at any time (but did not). To make the case even more like Joe's we can assume that, if he were to disable the cruise control, he could then choose any speed at which to travel.[51] What should we say about Tony regarding his speeding?

The thing that I think is important in this case is that the full explanation for why Tony was speeding won't merely be that he chose, as he came around the bend, to travel at 56mph. Rather, it will involve the speed being fixed at 56mph by the cruise control, him never attempting to disable the cruise control, and thus that condition never being unfixed. He didn't choose *despite* the cruise control being enabled. Instead, he passively

[51] One complication for this particular analogy is that, in a car, the act of disabling the cruise control setting itself results in slowing down. That is, the condition for unfixing is also a sufficient cause for a sort of choosing. We can ignore this detail, however, and assume that disabling the cruise control gives Tony the ability to choose a speed.

allowed the cruise control to guide the speed. This is what I would suggest is problematic in Joe's case as well—given his character and the presence of the buffer he's unable to actively choose his action in a way that allows for it to be up to him at that time. The kind of active choosing that confers responsibility, I would suggest, needs to involve things like consideration of reasons or deliberation. The issue is that allowing for a more active choice for Joe threatens to leave open alternative possibilities. Clarke was able to present a challenge to Strawson because he identified a place between reasons for actions and an action itself where libertarians can postulate undetermined, active choice. While Tax Evasion preserves indeterminacy in a general sense, Joe's character and the necessary-yet-unfulfilled condition fixes the ultimate action in such a way that there is no further choice. What I think ultimately explains this is that, even though the case preserves indeterminism in a strict sense, the action's being fixed plays the role of determinism in a responsibility-undermining way. This isn't because it eliminates alternative possibilities, but because it undermines the right kind of choice.

Another potential path towards responsibility around the time of action without deliberation comes from Pereboom himself. In a review of Kevin Timpe's *Free Will in Philosophical Theology*, Pereboom and his co-author Leigh Vicens discuss the intelligibility of libertarian action in the context of primal sin, especially as it concerns requests for contrastive explanations. The first scenario they discuss involves primal sin considering preferences in equipoise:

> Accordingly, we might imagine that Satan found himself, due to no fault of his own, with equally strong preferences for increased power and for the just outcome, but that he could have either rebelled or refrained from rebellion by his free will. Let's call this *the equipoise case*. Here the answer to the question, "Why did Satan choose to rebel?", is: "He wanted more power." To the question, "Why did he choose to

> rebel rather than not?", the answer is: "He just chose it." But…given libertarian commitments, it's natural to expect that there would be no other answer. (Pereboom & Vicens, 2015)

Here there's meant to be no unintelligibility in Satan's choice when considering either question posed. Initially, that he wanted more power provides an explanatory reason to the question of why he chose rebellion. If pressed for a contrastive explanation there won't be one that provides a *sufficient* cause, but this is standard on a libertarian account of free action. It may be unsatisfactory in a certain sense but there's no incoherence involved in an agent "just choosing" when preferences are in equipoise. The alternative scenario that Pereboom and Vicens consider involves Satan having a stronger preference towards sin:

> It's plausible, however, that in the paradigm case of sin the agent is not indifferent between sinning and not sinning, but instead *prefers* the expected outcome of the sinful action to that of the just action. Satan would then prefer his increased power to the just outcome -- and his motivations for rebelling and not rebelling would not be in equipoise. An option here is for the libertarian to agree that God created Satan with the preference to rebel or with a more general motivational source of this preference, but with the power to resist it in view of its injustice. So, then, in the moment of choice, Satan understands that rebelling is unjust, but he prefers it due to no fault of his own. He freely chooses to act in accord with his preference, a choice for which he is then blameworthy -- he could, after all, have avoided this choice by his free will. Let's call this *the sinful preference case*. (Pereboom & Vicens, 2015)

Once again there's no unintelligibility. On libertarian accounts we don't need, and in fact can't have, sufficient causes for action beyond the agent herself choosing (or, on an event-causal view, the occurrence of the set of agent-involving events relevant to the responsibility-conferring action). In this scenario we also get a contrastive explanation due to the stronger desire for increased power. Satan's choice, then, remains intelligible on either the equipoise or sinful preference case and he can be responsible for the

decision to rebel. My concern, of course, will be whether the considerations that Pereboom and Vicens discuss can be applied to the case of Joe in a way that secures a locus of responsibility at the time at which he evades taxes.

One thing to note first is that, with Joe, my claims aren't about "intelligibility" in the same way that it potentially arises in accounts of primal sin. The risk of unintelligibility with primal sin comes about because, in order for the sin to be primal, it must be that Satan decided to rebel in a responsibility-conferring way despite a set of preferences that he had no hand in generating. If the response to the contrastive question of why Satan chose sin over the just is that "he simply chose," then one might worry that the decision wasn't caused in a way that allows for responsibility. I'm willing to assume that this itself doesn't generate an underlying unintelligibility and, along with Pereboom and Vicens, that a decision like this can be intelligible and responsibility-conferring on a libertarian account. My suggestion of unintelligibility, if I'm indeed making one, doesn't involve the specific features of the way Joe's action comes about or whether he can be responsible given the case as a whole. It's not whether it makes sense for Joe to "just choose" given a set of preferences. Rather, it's that there's something wrong with calling Joe's taking the deduction a choice at all given the constraints imposed by the necessary-but-not-sufficient condition and the way in which the case proceeds. At the very least, I claim, it's not the kind of choice that secures the right locus of responsibility for a successful Frankfurt-style case. Still, it's worthwhile to take these two scenarios—equipoise and sinful preference—and consider how they may or may not map on to the case of Joe.

Regarding sinful preference, the case is essentially just like Tax Evasion but without

the features that make it a Frankfurt-style case. Pereboom created Joe with the preference to rebel and this explains, in a satisfactory way, why he ultimately takes the illegal deduction. Adding in the features of a buffered alternatives case doesn't change this, however it does change what we ought to say about his responsibility for taking the deduction.[52] Because he hasn't reached the requisite level of attentiveness, he can't deliberate about whether to take the deduction. If he's responsible at all, then, it's in virtue of being responsible for his preferences or an earlier, provisional decision to cheat on his taxes. At the time of the action it's simply not up to him in a way that would secure a new locus of responsibility. When considering the sinful preference case, then, my previous arguments will apply all the same. What, then, should be said about the equipoise case? First, we'll need to try to imagine Joe's case a bit differently. In particular, that rather than having a stronger preference for advancing his self-interest, his selfish desires and the desire to do the right thing are in equipoise. One immediate issue with this modification is that it becomes unclear what the necessary-but-not-sufficient condition is now doing. If Joe's preferences are truly in equipoise then why would we need *only* the buffer that prevents him from doing the right thing? It's hard to see how his competing desires could really be in equipoise if the buffer is asymmetric in this way since nothing could explain this asymmetry. This is a crucial question, since an asymmetric buffer is what secures the elimination of alternative possibilities. My initial concern, then, is that there is an inconsistency between Joe's libertarian free will, preferences that are in equipoise, and a buffer that only applies to one side of those

[52] The same would be the case, I'd argue, if we added the buffered alternatives features into the primal sin case.

preferences. Setting this aside, however, I still don't think that applying equipoised desires to Joe will allow for avoiding my dilemma. Unlike Pereboom and Vicens' interlocutors in the primal sin debate, I don't think that it's because "just choosing," or lack of a sufficient cause is itself unintelligible. Instead, I would suggest that this is not the *full* or *correct* explanation for why Joe ultimately cheats on his taxes. Even with his preferences in equipoise, it won't be the case that he "just chose" in the sense that it was up to him *what* he chose, or that "he could have either rebelled or refrained from rebellion by his free will," as was possible in the equipoised primal sin case. When asked why Joe cheated on his taxes the more appropriate explanation will be that "He *had* to—it was fixed at the time by his failing to meet the necessary-but-not-sufficient condition for doing right." Unlike just choosing, this explanation doesn't seem consistent with a libertarian free decision. We could press further and ask *why* he had to but that would eventually bring us back to the issue of explaining how the buffer is functioning as it does given preferences in equipoise. Developing Joe's case in this way still doesn't settle in a way that establishes responsibility at the time of action.

Finally, one key aspect of the discussion of primal sin involved the call for contrastive explanations and questions about whether that's appropriate given the nature of libertarian accounts of free will. This is worth addressing explicitly since, after all, in the course of presenting my arguments I've made (friendly) demands for contrastive explanations. For example, I claimed that there must be some explanation for why Joe, on his current path to action, will ultimately cheat on his taxes rather than the alternative. That is, I've asked what makes it the case that he's unlike Elaine, who will fail to cheat on her taxes unless she considers selfish reasons. There is a contrastive explanation for

this embedded in the case—that Joe has a strong inclination towards selfishness, which sets us up for a sinful preference-like case. My initial argument and further defense apply to this sort of case with no inappropriate call for contrastive explanation. Even if the only response is that "he just is that way," for example, it won't affect my claim that he doesn't secure a locus of responsibility around the time of action.[53] Adjusting the case to one in which Joe's preferences are in equipoise shifts my request from a contrastive explanation of his preferences to a contrastive explanation regarding the buffer more directly. We can then ask why Joe needs a "moral reason" condition rather than a "selfish reason" condition if his preferences are truly in equipoise. Even if this call for a contrastive explanation could be sidestepped, the inability to deliberate or reconsider his currently fixed path will prevent him from being directly responsible for the act of taking the deduction itself. My requests for contrastive explanations are *not* in relation to a choice to take the deduction and so are importantly unlike the request in the primal sin case. I agree that contrastive explanations are inappropriate when it comes to alternatives in a libertarian decision, though these alternatives are not what I request be contrasted.

2.5 The PAP, Leeway Theory, and Source Theory

If the dilemma I've posed for buffered alternatives cases is successful, then the PAP remains both a plausible control condition for blameworthiness and a viable understanding of the nature of free will. Traditional Frankfurt-style cases face the

[53] It may affect whether he's responsible *at all*, however. It could rule out that there is a prior locus of responsibility that's conferring responsibility for that later action.

traditional dilemma defense while other kinds of cases face their own unique challenges. Without buffered-alternatives cases to solidify the push against the PAP, it's intuitive force and connection to our ordinary blaming practices provide it sufficient credibility. Still, the failure of Frankfurt-style cases to present a genuine IRR scenario doesn't itself entail that leeway theory is the *right* account of control with respect to blameworthiness and free will. One could, as Frankfurt does, hold that even proposed IRR scenarios which don't meet the criteria for successful counterexamples demonstrate the truth of source theory.

In a piece reflecting on the debate regarding alternative possibilities, Frankfurt suggests that proposing a true counterexample to the PAP is not what was most important in his work. Rather, he claims that the importance of Frankfurt-style cases comes from something more general:

> The examples effectively undermine the appeal of PAP even if it is true that circumstances that do not bring an action about invariably leave open the possibility that the action might not be performed. What the examples are essentially intended to accomplish is to call attention to an important conceptual distinction. They are designed to show that making an action unavoidable *is not the same thing* as bringing it about that the action is performed. Their most pertinent import is that there is a difference between asserting that a set of circumstances possesses one of these features and asserting that it possesses the other. Appreciating this distinction tends to liberate us from the natural but nonetheless erroneous supposition that it is proper to regard people as morally responsible for what they have done only if they could have done otherwise. (Frankfurt, 2003, pp.339-40)

Regarding the PAP, according to Frankfurt, once you see how these cases go the veil of intuitive plausibility is lifted, despite sophisticated maneuvers to find alternative possibilities. We might think, then, that even if Frankfurt-style cases don't constitute IRR scenarios that they still reveal something significant about the nature of responsibility and control. That is, that what's *really* fundamental for attributions of

blameworthiness is being the appropriate source of one's actions. As someone who defends the importance of alternative possibilities I feel the weight of this claim, perhaps even more forcefully than the prospect of a genuine IRR scenario being developed. Challenging the specific construction of Frankfurt-style cases is crucial but to ignore the ultimate upshot in light of those challenges is to be evasive. The difficult question, then, is what conclusions should be drawn from Frankfurt-cases which are unsuccessful, at least in the sense that they don't eliminate alternative possibilities while preserving blameworthiness.

One possibility is simply that Frankfurt is right—even unsuccessful cases reveal that blameworthiness concerns being the appropriate source of one's actions rather than access to alternative possibilities. The problem with this conclusion is that it's hard to present an argument that both sides of the debate would find convincing. Source theorists might make the claim that it's just *made apparent* that Jones and Joe are blameworthy because they did as they wished and not because there were alternatives available. Leeway theorists can respond that the agents are blameworthy for doing as they wished *only because* they could have done something else instead. The argument points towards our intuitions about which condition for blameworthiness is being satisfied but those will mirror prior intuitions regarding the PAP. It would be the same if leeway theorists made the claim that failed Frankfurt-style cases establish the truth of the PAP. Absent a case that demonstrates that one set of intuitions is implausible, the result is a dialectical stalemate.[54] Another possibility is that Frankfurt is right, but with

[54] One might instead consider how a "neutral inquirer" would respond to the claim that Frankfurt-style cases demonstrate the truth of source theory even if they fail to constitute counterexamples to the PAP. This move was adopted by Pereboom (2008) and McKenna (2014) in relation to manipulation arguments,

a very important caveat. It could be that being the appropriate source of one's actions is the fundamental control condition for moral responsibility, as even failed Frankfurt-style cases reveal, *yet alternative possibilities are still necessary for blameworthiness*. Being the appropriate source of one's actions may itself require access to alternative possibilities. That is, there's something about deciding on one's own that necessarily involves deciding among alternatives. This appreciates the conceptual distinction that Frankfurt elucidates while offering an explanation for why, despite of this upshot, genuine IRR scenarios can't be developed. Of course, work would need to be done to articulate the precise connection between source and leeway principles. It would be a dialectically delicate endeavor as well, since it's inappropriate to *assume* that blameworthiness requires alternative possibilities. This is especially important given what's at stake in the debate over the PAP—the vast majority of compatibilists are committed to source views that can't also accommodate genuine metaphysical alternatives, even if they are less fundamental. However, if it's true that successful counterexamples to the PAP can't be developed, this itself may be reason to believe that alternative possibilities, in one role or another, are here to stay.

which rely significantly on intuitive judgments. It could be that a neutral inquirer would side with Frankfurt, though siding with the leeway theorist or failing to have firm intuitions in either direction don't seem unlikely.

References

Balaguer, Mark. (2010). *Free Will as an Open Scientific Problem*, Cambridge MA: MIT Press.

Chisholm, Roderick. (1964). Human Freedom and the Self. In *Free Will*, ed. Gary Watson, 26
37. Oxford: Oxford University Press.

Clarke, Randolph. (1993). "Toward A Credible Agent-Causal Account of Free Will." *Noûs* 27 (2): 191–203.

Clarke, Randolph. (2005). On an Argument for the Impossibility of Moral Responsibility. *Midwest Studies in Philosophy, 29*, 13–24.

Fischer, John Martin. (1994). *The Metaphysics of Free Will*. Oxford, Blackwell Publishers

Fischer, John Martin, and Neal A. Tognazzini. (2009). "The Truth about Tracing." *Nous* 43 (3): 531–56.

Frankfurt, Harry. (1969). Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy* 66 (23): 829–39.

Frankfurt, Harry. (1969). Some Thoughts Concerning PAP. 339-345. In *Moral Responsibility and Alternative Possibilities*. Ed. by Michael McKenna & David Widerker. Aldershot, U.K.: Ashgate.

Ginet, Carl. (1966). Might We Have No Choice? In *Freedom and Determinism*, ed. Keith Lehrer, 87–104. New York: Random House.

Ginet, Carl. (1996). In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing. *Philosophical Perspectives, 10*, 403–417.

Ginet, Carl. (1997). Freedom, Responsibility, and Agency. *The Journal of Ethics* 1.1: 85-98.

Goetz, Stewart (1999) Stumping for Widerker, *Faith and Philosophy: Journal of the Society of Christian Philosophers*: Vol. 16: Iss. 1, Article 6.

Haji, Ishtiyaque, & McKenna, Michael. (2004). Dialectical Delicacies in the Debate about Freedom and Alternative Possibilities. *The Journal of Philosophy*, *101*(6), 299–314.

Hunt, David P. (2005). Moral Responsibility and Buffered Alternatives. *Midwest Studies in Philosophy*, *29*, 126–145.

Hunt, David, & Shabo, Seth. (2013). Frankfurt cases and the (in)significance of timing: a defense of the buffering strategy. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *164*(3), 599–622.

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

Kane, Robert. (1998). *The Significance of Free Will.* New York: Oxford University Press.

McKenna, Michael. (2008). Frankfurt's Argument against Alternative Possibilities: Looking beyond the Examples. *Noûs*, *42*(4), 770–793.

McKenna, Michael. (2008b). Ultimacy and Sweet Jane. 187-208. In *Essays on Free will and Moral Responsibility* Ed. by Trakakis, Nick and Cohen, Daniel. *Reference & Research Book News*, *25*(1).

McKenna, Michael. (2008c). A Hard-line Reply to Pereboom's Four-Case Manipulation Argument. *Philosophy and Phenomenological Research*, *77*(1).

McKenna, Michael. (2014). Resisting the Manipulation Argument: A Hard-Liner Takes It on the Chin. *Philosophy and Phenomenological Research*, *89*(2).

Mele, Alfred R. (1995). *Autonomous Agents: From Self-Control to Autonomy*, New York: Oxford University Press.

Mele, Alfred R., & Robb, David. (1998). Rescuing Frankfurt-Style Cases. *The Philosophical Review*, *107*(1), 97–112.

Mele, Alfred R., & Robb, David. (2003). BBs, Magnets and Seesaws: The Metaphysics of Frankfurt Style Cases. 128-138. In *Moral Responsibility and Alternative Possibilities*. Ed. by Michael McKenna & David Widerker. Aldershot, U.K.: Ashgate

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

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

Pereboom, Derk. (2001). *Living Without Free Will*. Cambridge University Press.

Pereboom, Derk. (2008). A Hard-line Reply to the Multiple-Case Manipulation Argument. *Philosophy and Phenomenological Research*, *77*(1)

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

Pereboom, Derk, & Vicens, Leigh. (2015). Review of Kevin Timpe, Free Will in Philosophical Theology. *Notre Dame Philosophical Reviews*. Online. (ndpr.nd.edu).

Schaffer, J. (2000). Trumping Preemption. *The Journal of Philosophy*, *97*(4), 165–181.

Strawson, Galen. (1994). The Impossibility of Moral Responsibility. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *75*(1/2), 5–24.

Stump, Eleonore. (1996). Libertarian Freedom and the Principle of Alternative Possibilities. 73-88. In *Faith, Freedom, and Rationality* Ed. by Jordan, Jeffand Snyder, Daniel Howard Lanham MD: Rowman and Littlefield.

van Inwagen, Peter. (1975). The Incompatibility of Free Will and Determinism. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *27*(3), 185–199

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

Vargas, Manuel. (2005). "The Trouble with Tracing." *Midwest Studies in Philosophy* 29 (1): 269–91.

Vihvelin, Kadri. (2004). Free Will Demystified: A Dispositional Account. *Philosophical Topics*, *32*(1/2), 427–450.

Widerker, David. (1995). Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities. *The Philosophical Review*, *104*(2), 247–261.

Widerker, David. (2000). Frankfurt's Attack on the Principle of Alternative Possibilities: A Further Look. *Philosophical Perspectives*, *14*, 181–201.

Wiggins, David. (1973). "Towards a Reasonable Libertarianism." In *Essays on Freedom and Action*, ed. Ted Honderich, 31–62. London: Routledge and Kegan Paul.

Wolf, Susan. (1980). Asymmetrical Freedom. *The Journal of Philosophy*, *77*(3), 151–166.

CHAPTER 3

PRAISEWORTHINESS AND REASONS FOR ACTION

3.1 Introduction

Discussions of praiseworthiness arise in two relatively separate philosophical contexts—that of moral responsibility, on the one hand, and that of moral worth, on the other. While the background and perhaps even the ultimate motivation may be different in each context, both get at the same fundamental question. That is, whether and when it's appropriate to praise agents for particular actions. The moral responsibility literature focuses more directly on the praiseworthiness of agents while questions about moral worth concern the status of actions, but these come together in a way that allows for treating both at once. Following Nomy Arpaly, we can use phrases like 'a morally praiseworthy action' and 'an action with positive moral worth' interchangeably (Arpaly, 2002, Ch.3 p.4). This makes the aspect of praise explicit when discussing the moral worth of actions, but we can also spell out the connection to the assessment of agents. For example, Julia Markovits characterizes morally worthy actions as "actions for which the agent who performs them *merits praise*" (Markovits, 2010, p.203). Accordingly, we can take an agent to be praiseworthy if and only if she's performed a praiseworthy action. This needs many clarifications, of course (a task to which I dedicate the next section). One immediate caveat though is that the actions in question must have *moral* worth in order to merit *moral* praise. As Markovits points out, certain actions may warrant praise due to non-moral considerations (Markovits, 2010, p.203).

These may include demonstrations of skills and abilities like athleticism, artistry, or even certain kinds of intelligence. Whether it's appropriate to praise in these cases may depend on the extent to which the ability is cultivated rather than innate.[55] Regardless, while these types of actions might warrant a kind of praise, it won't be in virtue of the moral worth of the action and thus the agent won't be *morally* praiseworthy.

What's interesting about these two philosophical contexts is that a similar account of praiseworthiness is prominent in both. Namely, that being praiseworthy involves *doing the right thing for the right reason*. Proponents of such a view in the literature on moral responsibility include Susan Wolf (1980) and Dana Nelkin (2008, 2011), while both Markovits (2010) and Arpaly (2002) defend the view in the moral worth literature. These accounts will differ on the details, and even how much detail is provided, but the overarching sentiment will be the same—praiseworthiness is intimately tied to reasons for action and, in particular, acting on the *right* reasons. My aim in what follows is to consider how *right thing for the right reason* (RTRR) accounts of praiseworthiness fair against potential problem cases. That is, what they say or ought to say about certain challenging cases and whether that accords with ordinary intuitions about praiseworthiness. In order to do this I'll lay out several kinds of cases, varied based on the action or the agent's reasons for action, and consider each in turn. I'll begin, however, with some preliminary work—in section 2 I'll clarify how I'm understanding praise for the purposes of this discussion and in section 3 I'll sketch the kinds of case structures I'll be considering. Then, section 4 will include a more detailed presentation

[55] Of course, some may find it intuitive that praise is just as appropriate when a skill or ability is innate. In fact, it's not unusual for praise to be expressed for innate physical attributes like height or eye color. Since my concern will be with moral praise, I'll set these considerations aside.

of RTRR accounts, as well as how they apply in cases where an agent acts for the wrong reason. In sections 5, 6, and 7 I'll consider certain cases that, at least initially, seem hard to capture on RTRR accounts. Finally, in section 8 I'll discuss what I think we can take away from consideration of these kinds of cases.

3.2 Praiseworthiness

As we've seen already, it may be appropriate to praise some actions or agents in non-moral contexts. Even setting that aside, however, there are still important clarifications to make regarding how I'll be understanding praiseworthiness. First, I intend to focus only on praiseworthiness for particular actions and not praiseworthiness of character. While most would agree that certain people have dispositions, or even entire lives which are morally praiseworthy, this is often or always in virtue of having performed praiseworthy actions to a sufficient degree. Of course, it may be possible for a person to have praiseworthy dispositions or live a praiseworthy life without having been praiseworthy for any particular action. Consistently resisting temptation, or simply failing to perform any (or many) blameworthy actions may be sufficient for praiseworthiness of character. There may also be broader reasons for wanting to cultivate this kind of character, but my discussion will center around reasons, or lack of reasons, for performing actions.[56] Next, the sort of praise I have in mind is *basically deserved* praise. I assume that some agents perform actions for which they are

[56] As will become clear, I'll include omissions, such as 'refraining to do *x*,' in the broader class of "actions." Going forward, I'll use the term "action" to mean either an active action or an omission (though I hope the context will be clear in each instance).

responsible merely for having performed the relevant action and that, in some cases, these agents basically deserve praise. This declaration is often seen in the moral responsibility literature and so one thing to note is that "merely for having performed the relevant action" is not meant to say anything substantive about what makes an action praiseworthy. That is, reference to the *mere performance* of an action should not be taken to be sufficient for responsibility. Rather, the emphasis is that there are no other considerations by which we're basing responsibility aside from having performed an action with moral worth. We can contrast this with a consequentialist, or forward-looking sort of responsibility in which praising or blaming is appropriate as a means of moral formation for the target, deterring or encouraging an agent to repeat similar actions in the future, or (more controversially) deterring or encouraging the behavior in others. Whether responsibility in this latter, forward looking sense is appropriate will not necessarily turn on whether an agent had reasons to perform an action or its alternative and so it won't be my concern here. Rather, my focus will be praise in the former sense, which characteristically involves the appropriateness of positive reactive attitudes like expressions of approbation and gratitude (see, e.g., Strawson, 1962). In taking an agent to be praiseworthy if and only if she's performed a praiseworthy action, then, we can think of her as being *deserving*, *meriting*, or *worthy* of moral praise regardless of whether it would be better overall not to praise her given other circumstances.

Finally, there will certainly be disagreement about whether and to what extent particular actions are praiseworthy. The right thing in a given situation will sometimes be unclear to almost everyone or two people may each have very clear but opposing

judgments. Thoughts about praiseworthiness may vary with intuitions or commitments to different first-order ethical theories. Volunteering to serve in a war, for example, may be seen as honorable by some but objectionable to others. Taking care of one's children can be viewed as noble, and thus praiseworthy, but some may view it as merely meeting expectations, and thus neutral. How one feels about these cases will depend on intuitions, background views about what is moral and immoral, or how moral obligations are understood. Judgments about praiseworthiness will also vary depending on the context in which the action is performed. Allowing a friend to finish off your bottle of water is a cordial act when performed at the park but becomes something much more admirable when walking through the desert. Moreover, even when an action is agreed to be morally good, what the *right* reason for action is in a particular case may not be obvious and judgments may differ. For example, one person might think that the only right reason not to eat meat has to do with the suffering of animals. Another person may not be moved by concerns for animal welfare but think that we shouldn't eat meat because of the environmental impact of factory farming. I'll attempt to avoid these issues by centering the discussion around what I hope are fairly uncontroversial cases. Moreover, I expect that the details of any particular case will not be crucial for generating an interesting discussion regarding praiseworthiness on RTRR accounts. Rather, the important feature of these cases will be their structure, within which readers may substitute their own preferred cases if necessary.

3.3 Structuring Cases of Potential Praiseworthiness

Having clarified the sort of praiseworthiness I have in mind I'll now turn to the types

of cases I will (and won't) be concerned with. First, independent of any particular theory, there will be many uncontroversial cases of agents deserving praise for their actions. When someone performs some a good action and the motivation for the action is good, often they will very clearly deserve praise. A couple who adopts a child in need, for example, would be considered praiseworthy for that action if their motivation was to, say, relieve the child's suffering or give her the best life possible. This intuition could be made even stronger if the adoption involved a great degree of sacrifice on the part of the couple. Whether or not agents like these are praiseworthy is uncontroversial, at least when it comes to the question of how motivation relates to praiseworthiness.[57] Doing the right thing with the best of intentions (broadly speaking) is the kind of case that any reasonable account of praiseworthiness will capture. There is, however, another kind of case that fits the above schema but is not as clear-cut. We can imagine that an agent performs an action that is itself good, and so potentially praiseworthy, but that the motivation for acting is not itself something that's commendable. In fact, the impetus for acting could be downright nefarious, even if the action itself usually merits praise. Here we have cases of *doing right for the wrong reason*. Amending the previous example, we might imagine a couple who adopts a child in need, not to relieve suffering or concern for providing a good life, but because the child will be a source of cheap labor at the family restaurant. It becomes a little harder to say whether this initial action (adopting a child in need) is praiseworthy, especially if other plausible assumptions are made (e.g., that the child would be better off overall).

[57] I'm here, and for the remainder of the discussion, setting aside any broad basic-desert skepticism (which would result in any purported case of deserved praise being controversial).

Other cases worth considering involve praiseworthy omissions. Here again there will be uncontroversial cases which any account of praiseworthiness should (and I think will) be able to explain. These would include cases where an action is morally wrong but the agent refrains even though she has strong motivation for performing the action. An example here is a mother who, struggling to feed her kids, could reliably steal twenty dollars per week from a coworker with whom she shares an office. Whether this would truly be a morally wrong action may depend on the details of the case, though we can assume for the present purposes that it would be immoral. We can also assume that she has strong reasons to steal, is conscious of those reasons, and if she decided to steal she could do so undetected and indefinitely. Still, she ultimately refrains on moral grounds. Now, resisting great temptation, especially with no risk of blame or punishment, will certainly be considered praiseworthy among many people. [58] Accordingly, this should be another uncontroversial case as it concerns motivation for action. This becomes even more plausible when considering actions with positive moral worth to be *virtuous* actions. As Philippa Foot outlines, virtues are corrective in that "each one [stands] at a point at which there is some temptation to be resisted or deficiency of motivation to be made good" (Foot, 2002, Ch.1 p.9). Considering virtuous *actions*, then, we can view at least some of them as those performed by an agent while tempted by, and resisting, some vice. With omissions, though, there are at least two other types of cases that leave room for interesting reasons-related discussion. Similar to *doing right for the wrong reason*, there may also be situations that involve *resisting wrongdoing for the wrong*

[58] That is, at least, no risk of blame or punishment from *others*. Whether people in this situation would or should have feelings of guilt and remorse upon stealing the money is another issue.

reason. As before, there will be something about the agent motivating the morally correct (and typically praiseworthy) action, but the motivation won't itself be commendable. An example of this sort is a drug kingpin, who watches a bag of cash fall off of a truck which then disappears into the night. He may refrain from taking the cash, and even report the money to the appropriate institution, but not for any noble reason. Instead, he may know that the police are raiding his home in the morning, he has no place to stash the money in the meantime, and that when it's found it will result in more criminal charges than he was already facing. The action itself might be morally good, and even praiseworthy under more ordinary circumstances, but the reasons for action may make it so that this agent does not deserve praise.

Another interesting case of praiseworthy omission involves refraining from performing a wrong action when the agent isn't motivated to perform that action at all. Here we can imagine a young woman who only needs a good score on her college entrance exam in order to be admitted to a prestigious university. As luck would have it, she comes across the answers to the exam and could quite easily memorize most of them beforehand and no one would ever know. She refrains from looking at the answers, not on moral grounds but because there is nothing about her motivations that would be furthered by her succeeding on the exam, and thus nothing furthered by her cheating. We can suppose that it would not be a bad thing if she were to do well on the exam, and were admitted to a good university, but that it wouldn't help (or hurt) any of her life's ambitions. I'll call these situations *refraining from wrongdoing with indifference*. Of the three types of cases I've outlined so far, it's clear that the first two are more similar than the last. Accordingly, I'll treat the first two together under the label of *right outcome,*

wrong reason and discuss *refraining from wrongdoing with indifference* in isolation. In particular, I'll be considering how RTRR accounts handle these kinds of cases. Afterwards I'll discuss two additional kinds of cases—*doing right with indifference* and *doing the suboptimal for the right reason*—in the context of RTRR accounts.

3.4 RTRR and Right Outcome, Wrong Reason

To begin discussion of *right outcome, wrong reason* cases I'll return to the adoption example—a couple adopts a child in need, not for good moral reasons but because the child will be a source of cheap labor at their restaurant.[59] What should we say about this case? First, the action itself is the sort that would merit praise under more typical circumstances. People who adopt children often deserve praise since (I suspect and hope) it's usually done for good moral reasons. Next, we can make assumptions about the case that makes the "right outcome" aspect more plausible. The child may be better off with the adoptive parents than she would be remaining in her current situation. After all, simply being a child in the foster care system can take a psychological toll, especially as one gets older, let alone a heightened risk of mistreatment. Moreover, we can assume that there is no alternative where the child would be better off and so this is the *best* outcome for her. Finally, we can assume that the couple willingly performed the action that resulted in the best outcome. Even with these considerations in mind, though, this couple just doesn't intuitively deserve praise.

[59] I'll be setting aside any issues of shared or distributed responsibility here. When asking whether *they* are responsible I'll simply mean either of them, or each individually.

It seems, then, that reasons and motivation may be playing the key role in whether or not these agents are praiseworthy. In particular, that in order to be praiseworthy for a particular action it's not enough merely to perform the right action. Rather, praiseworthiness may require the further conditions stipulated by a RTRR account. In the moral responsibility literature both Wolf and Nelkin have put forth versions of RTRR accounts. In defending an asymmetric view of the conditions of blame and praise, Wolf speaks of praise in terms of "doing the right thing for just the right reason" (Wolf, 1980, p.156). What's meant to come out of this view is that, while blameworthiness requires the ability to do otherwise, praiseworthiness does not. Praiseworthiness will involve aiming at what she calls "the True and the Good" and need not require having had the ability to aim at something else (Wolf, 1980, p.160). Nelkin also defends the view that responsibility is asymmetric in this way. According to Nelkin, "people are responsible when they act with the ability to do the right thing for the right reasons, or a good thing for good reasons" (Nelkin, 2008, p.497). The second clause here is meant to capture cases where there is no *single* right thing to do, or no *one* right reason for action. For example, we could imagine that a person has the opportunity to volunteer at their local library or a nearby animal shelter, doing either of which would be equally morally good. Moreover, if the person picks the animal shelter they might be motivated to do so because they want to ensure the animals have the best care while in the shelter or because they want to aid in finding the animals new homes. Either of these reasons seems like a good reason and, in particular, a good enough reason to justify praise.[60]

[60] Going forward I'll use "the right reason" and "the right reasons" interchangeably, except in cases where the distinction becomes important.

The upshot of these two accounts from the moral responsibility literature is meant to be that praise, perhaps unlike blame, does not require the ability to have performed an alternative action—neither genuine metaphysical alternatives nor considered counterfactually. That is, agents can merit praise simply for having performed the right thing for the right reason. In the literature on moral worth, however, we get more detailed accounts of what it *means* to do the right thing for the right reason.

One RTRR account in the moral worth literature comes from Julia Markovits and can be captured in what she calls the "Coincident Reasons Thesis" (CRT):

> *My action is morally worthy if and only if my motivating reasons for acting coincide with the reasons morally justifying the action*—that is, if and only if I perform the action I morally ought to perform, for the (normative) reasons why it morally ought to be performed. My motivating reason for performing some action in this case will not be the duty-based reason "that the moral law requires it" but the reasons *for which* the moral law requires it. (Markovits, 2010, p.205, emphasis in original)

Two things are worth emphasizing about this characterization. First, the thesis is stated in terms of *moral worth* rather than praiseworthiness directly. As noted at the outset though, we can take actions with positive moral worth to be actions for which the agent deserves praise. Second, for an action to have positive moral worth it's neither necessary nor sufficient, as it was for Kant, that it be done merely from "the motive of duty."[61] In fact, someone performing an action solely because duty requires it will not be praiseworthy at all on this account. Instead, on Markovits' account the motivating reason needs to be the reason *why* duty requires a particular action. A useful distinction here is consideration of doing the right thing *de dicto* vs *de re*. That is, it's possible that

[61] For a defense of Kant's commitment to these claims see, e.g., Stratton-Lake (2000).

an agent performs a right action, motivated by the desire *to do the right thing*, whatever the right thing turns out to be. This won't be enough (or required) for praiseworthiness on Markovits' CRT. Rather, what's necessary and sufficient for praiseworthiness is that the agent be motivated by the right reason *de re*, or the actual justifying reason that fits the description 'the right reason.'

Another RTRR account of positive moral worth (and thus praiseworthiness) comes from Arpaly and is dubbed "Praiseworthiness as Responsiveness to Moral Reasons" (PRMR):

> For an agent to be morally praiseworthy for doing the right thing is for her to have done the right thing for the relevant moral reasons—that is, for the reasons for which the action is right (the right reasons clause); and an agent is more praiseworthy, other things being equal, the deeper the moral concern that has led to her action (the concern clause). Moral concern is to be understood as concern for what is in fact morally relevant and not as concern for what the agent takes to be morally relevant. (Arpaly, 2002, Ch.3 p.19)

As we can see, Arpaly also rejects the idea that praiseworthiness involves acting from the motive of duty. On her view, like on Markovits', "For a right action to have (positive) moral worth, it is neither sufficient nor necessary that it stem from the agent's interest in the rightness of his action" (Arpaly, 2002, Ch.3 p.8). The lack of sufficiency comes because one can do the right thing by accident while being mistaken about what the relevant moral reasons are, even when they may be acting because they want to do right. Intuitively, praise appears inappropriate in such cases. The lack of necessity is illustrated by cases where one does the right thing, motivated by the reasons for which

the action is right, while unconcerned by or acting contrary to duty.[62] One difference between Markovits and Arpaly's RTRR accounts is that, in addition to acting for the right reasons, moral concern also plays a key role in praiseworthiness. That is, the degree of moral concern an agent acts with will determine the degree to which she's praiseworthy. Concern should not be understood as intensity of feeling or commitment according to Arpaly. Rather, concern is associated with motivational strength, emotional investment, and being "morally conscious" (Arpaly, 2002, Ch.3 pp.20-2). While Markovits doesn't incorporate moral concern, she does provide an account of how to understand degrees of praiseworthiness. On her view, the degree of moral worth of actions varies "to the degree that the reasons motivating them coincide with the reasons morally justifying them" (Markovits, 2010, p.237).[63] When discussing RTRR accounts I'll have both of Markovits' CRT and Arpaly's PRMR in mind while noting, in select places, where they may differ. I'll set aside the RTRR accounts from the moral responsibility literature for the now, though I will return to them briefly in section 8.

RTRR accounts of the sort Markovits and Arpaly outline seems to explain the difference between the two adoption examples I mentioned earlier. That is, why the couple who adopts with the best intentions deserves praise while the couple who adopts for a source of cheap labor does not. I believe most will share my intuition that the drug kingpin doesn't deserve praise either and the fact that he failed to do the right thing for

[62] Mark Twain's *Adventures of Huckleberry Finn* contains a commonly used example of this kind of case (see, e.g., Arpaly, 2002, Ch.3, p.10; Markovits, 2010, p.208). Huck Finn, believing that it's wrong to help a slave escape, does so anyway, presumably for the actual justifying reasons. As a result, he may be praiseworthy despite acting *against* what he believes is the motive of duty.

[63] Nelkin has also defended the view that praiseworthiness comes in degrees. On her view, the degree of praiseworthiness varies according to difficulty, understood in terms of effort and sacrifice (Nelkin, 2016).

the right reason explains this as well. Rather than reporting the money for the right reason—so that it finds its rightful owner, perhaps—he reports the money out of concern for his own self-interest. Neither the couple nor the kingpin acts for the right reason and, moreover, we might think that had their actual selfish reasons for doing the right thing not been in place they would not have done the right thing at all.[64] As we would expect, then, RTRR accounts handle straightforward *right outcome wrong reason* cases very well. A concern, however, is that these accounts may have trouble correctly diagnosing other cases. The cases I've labeled *refraining from wrongdoing with indifference, doing right with indifference,* and *doing the suboptimal for the right reasons* will be among these, all of which I'll discuss in the following two sections. Before that, however, it's worth considering a variant of a *right outcome wrong reason* case that's less straightforward.

Beyond agents doing the right thing for a bad reason (as was the case with the selfish adopting couple) or for, at the very least, no good reason (like the drug kingpin), it's possible for an agent to do the right thing for a good reason if still not the best reason. That is, we can imagine an agent who does the right thing, perhaps not for *the* right reason, but for some other commendable reason instead. An example of this kind is a man who, upon seeing a stranger drowning in a lake, swims out to save the stranger. Now, the right reason in this instance may be something like preserving life, respecting the stranger's humanity, or preventing suffering. We can suppose that none of *these*

[64] This is not necessarily the case, of course. We could imagine cases where the couple would adopt anyway, or the kingpin would report the money, and that either of these actions would have been performed for good reasons had the selfish reasons not been in place.

reasons motivated the man and instead he swims out because his kids are with him and he wants to teach them to do the right thing. I take it that contributing positively to the moral formation of one's children is a commendable motivation, and so the man has performed the right action for *a* good reason, though unmotivated by the *right* reason. The man's psychology may be important for determining whether he's intuitively praiseworthy and the case can be built in different ways to make a judgment that he may be praiseworthy more plausible. First, he could know what the right thing to do is, he could know the right reason (but not be motivated by it), and thus is looking to teach his children both. In order to make this kind of agent more plausible psychologically we could assume that *in general* he cares about the right-making reasons for this kind of action, just not *today*. He may be too overwhelmed, or feeling depressed, but he still cares too much about his children to pass up an opportunity to demonstrate what's right. Alternatively, he may know what the right thing is, not be cognizant of the right reason, and thus only seek to show them an example of doing the right thing. Without a good explanation for the action this may not be the ideal moral lesson for his children, but at least he's getting them halfway there (and, due to his ignorance, maybe that's the best he can do at the present time). I'll continue with the first understanding of his psychology since it seems much more plausible, although I do think that an agent with the second kind of psychology is at least conceivable.[65] There will be two important questions regarding a case like this—first, what RTRR accounts would say about

[65] In order to motivate an intuition of praiseworthiness on this second understanding, however, we may have to also make clear that his ignorance about the right reason for saving a stranger is itself justified or blameless rather than willful.

whether the man is praiseworthy and, second, what they intuitively ought to say.

When it comes to what RTRR accounts ought to say I suspect that intuitions are less clear than they are regarding the selfish couple and the kingpin. My own initial reaction is that he's not praiseworthy and this has something to do with his indifference to the life of the stranger drowning (or, at least, the fact that he doesn't care non-instrumentally). However, I think there are ways to imagine the case which emphasize that he's not simply callous and uncaring. He could instead fail to be motivated by the right-making reason in a more sympathetic way. Taking the psychological features mentioned above, he could be prone to depression and in this situation that depression overwhelms the motivation he may otherwise have had to adequately consider the life of the stranger. Still, he loves his children so much that swimming out to save the stranger for *their* benefit is enough motivation to act. Thinking about the case in this light makes it much more plausible, at least to me, that he's praiseworthy for the act of saving the stranger. Not only did he do the right thing but it was at great personal risk and he did it for a good reason. Adding to this that he seems to have a good excuse for not being motivated by *the* right reason at the time may be enough to swing institutions. Regarding what RTRR accounts would have to say about this case, the immediate verdict would seem to be that the man is not praiseworthy. Although he's done the right thing for a good reason, he hasn't done the right thing for *the* right reason. Even taking Nelkin's statement that one could also be praiseworthy for doing "a good thing for good reasons," which Markovits and Arpaly would also endorse, it doesn't seem to help. Allowing for praiseworthiness under these circumstances is meant to capture cases where there are multiple, equally right justifying reasons and not cases where the agent

fails to act for the real, obvious right-making reason. Even if there are many right reasons for saving the strangers life it doesn't seem plausible that "instilling moral values in one's children" would be among them. It would appear, then, that RTRR accounts are forced to say that the man is not praiseworthy. Whether we think that's a bad thing, or a counterexample, will depend on our intuitions about the case. However, there is something that defenders of RTRR accounts could say to make the judgement that he's not praiseworthy more palatable. That is, even if he's not praiseworthy for saving the stranger's life he's still demonstrated a virtuous character in a certain sense. Being so committed to his children is a feature of his character that merits praise, and one that's worthy of admiration.

3.5 Refraining from Wrongdoing with Indifference

It's clear that one particular feature of both the selfish adoption case and the kingpin case are guiding intuitions about praiseworthiness. In each case, the agents not only fail to act for the right reason but act for some off-putting reason instead. For the selfish couple, in fact, their reasons for action not only fail to justify praise but may result in their deserving *blame* for their action. Using a child as a means in this way, regardless of whether it improves the child's life, may be blameworthy. The kingpin may not deserve blame for acting in his own self-interest but doing the right thing merely out of selfishness is certainly distasteful. Initially, without our more sympathetic assumptions, the case of the man who saves the stranger in the lake can also be thought to have this feature (though perhaps in a less obvious way). He performs the right action for a good

reason but the fact that he isn't concerned about the stranger's life evokes negative attitudes. Even if he acted for *a* good reason, he's not motivated by *the* right reason, which in this particular case is one that every moral agent should be motivated by (whether or not they can bring themselves to act on it in these dire circumstances).[66] Lacking such a basic concern for the well-being of others would make this character offensive, to say the least. In each of these cases it's not merely that the agents don't deserve praise, but it would feel *wrong* to praise them. A better, or perhaps more complete test of RTRR accounts will need to involve cases where the right reason is missing but neither the motivating reason nor the absence of the right reason is objectionable. I take the clarified version of my stranger drowning case to have these features, but there are others still.

What then of cases of *refraining from wrongdoing with indifference*? For considering cases with this structure we can return to the example of the young woman who, although she has access to the answers for a college entrance exam, she doesn't use them to help herself get admitted to a prestigious university. In this case we can assume that there's nothing blameworthy, or even off-putting about her reasons for refraining from cheating. Rather, there is simply nothing in her motivational set that would be furthered by succeeding on the exam, and thus nothing furthered by using the answers. Perhaps she's indifferent between staying where she is and moving to the new city where the university is located and so taking the exam is a way of choosing among

[66] Markovits' own internalist account of reasons has the consequence that all agents *do* have reason to have this sort of concern whether they acknowledge it or not, given that they will anything at all (Markovits, 2014).

equally preferable alternatives. Like with the other cases, I believe the action itself is praiseworthy under more typical circumstances—when applicants have a great desire to attend a university, and many of their life goals may depend on them doing so, resisting the temptation to cheat undetected is admirable. We can also assume that she considers but decides not to use the answers, and so the typically praiseworthy action was performed willingly.

Although there isn't anything that feels *wrong* about praising the young woman, as she has no objectionable motivations, it still doesn't seem intuitive that she *deserves* praise. This may be because she did the right thing but not for the right reason, but there's another possible explanation as well. That is, perhaps what elicits the intuition that she's not praiseworthy is that there was no *temptation* to cheat given her goals. She's not motivated to do better on the exam than she would do taking it on her own merits and so wouldn't give cheating much consideration. If resisting temptation is essential to being praiseworthy for an omission, then this could explain why the young woman doesn't deserve praise. This could be thought to explain her lack of praiseworthiness *instead* of RTRR considerations, or suggest that such accounts need to incorporate resisting temptation when it comes to praiseworthy omissions. However, I don't think that the fact that she's not tempted to cheat is what's doing the work here.[67] First, as we've already seen, doing the right thing while resisting temptation is not sufficient for an action being praiseworthy. The kingpin reported the lost money and we can assume that this was a difficult task—perhaps the thought of passing up the

[67]At least not directly, that is. I'll return to this question in section 8 and suggest that perhaps it is playing a more indirect role in a case like this.

opportunity to take it agonized him, even knowing that in reality it would be of no use. His temptation was outweighed by other factors but it was still temptation. The kingpin then, resisted wrongdoing but is not praiseworthy for his action. I think the case can be made that resisting temptation is not necessary for praiseworthiness either. To illustrate this we can imagine a nurse with a sick child at home who would benefit greatly from a scarce and expensive medication. The nurse may come across some of this medication at the hospital, be able to pocket it unnoticed, but doesn't do so. She refrains simply because she's deeply committed to her moral values and, in fact, wouldn't even give stealing the medication any real consideration. She realizes that stealing is wrong and, perhaps more importantly given an RTRR account, that others may be entitled to this very dose. This is despite the fact that she has great personal reasons for taking it—improving the health of her child. Although the nurse does not resist temptation I believe she can be praiseworthy for her action. What this case, along with that of the kingpin, seems to turn on then is not whether they resisted temptation but rather whether they performed the right action for the right reason. The kingpin was tempted but failed in this regard, and so is not praiseworthy. The nurse was not tempted but acted for the right reason and is still praiseworthy.

Removing the lack of temptation as an explanation for why the young woman doesn't deserve praise shifts the focus back to her reasons for action. Since her actual motivations aren't objectionable in any way (unlike those of the selfish couple and the kingpin) it can't be said that this is what shapes judgments about the case. Instead, the best explanation seems to be that she's not praiseworthy because she wasn't motivated by the right-making reasons. It could be that had she been more disposed to do well on

the exam she still wouldn't have cheated, and that she would have refrained for the right reason, but this would be more like a straightforward case of praiseworthy omission. Even those with praiseworthy characters can perform actions which are typically praiseworthy, but not praiseworthy in some particular situation. In the circumstances outlined there simply isn't the right connection between the good action and the young woman's motivations.

3.6 Doing Right with Indifference

While the young woman refraining from cheating with indifference doesn't warrant praise and RTRR accounts plausibly explains this, there are two more cases I'll discuss. First, another type of "indifference" case involves *doing right with indifference*. Here we can consider an agent who, rather than refraining from wrongdoing, actively performs a typically praiseworthy action with indifference. An example of this kind of case is a retired man who, upon leaving his career, decides to volunteer as a crossing guard for the local elementary school. We can imagine that he was motivated, at least initially, by his concern for the safety of the children. Over the years, however, waking up early each day to ensure the children's safe crossing has simply become "what he does." His motivation for heading to the crosswalk is no longer the well-being of the children. Is the retiree praiseworthy, even after his good action has become simply part of the routine? The answer from the perspective of an RTRR account may be that it depends.

Even if initially he was motivated by the right justifying reason, and thus was

praiseworthy early on in his volunteering, the retiree may no longer deserve praise. He may now be acting *merely* out of habit, not motivated at all by, say, the safety of the children. In this case it seems that he wouldn't be acting for a good reason at all.[68] On this understanding of the case it has an important similarity with that of the young woman who refrains from cheating on her exam. That is, along with a disinterest in the particular right-making reasons, the retiree is disinterested in any moral features of the action. Intuitively, there's nothing that praiseworthiness would attach to aside from the mere performance of a good action, which seems implausible. We could also consider a version where, rather than acting merely out of habit, he's still interested in the morality of his action in another way. It could be that, although he's no longer particularly concerned about the children's well-being, it's become "his duty" to help them cross the street each morning. This is what motivates him to continue doing the right thing each day even though he remains disinterested in what might make that a duty in general.[69] Of course, one thing to note is that this is now a case of doing right only *with a certain kind of indifference*. He won't be indifferent to the moral relevance of the action completely as we imagined before, but he'll remain indifferent to the particular right-making reasons for the action. As we would expect, though, this is just the sort of indifference that would concern the two RTRR accounts I've focused on.

On either Markovits' CRT or Arpaly's PRMR, the fact that the retiree is acting from

[68] We may be tempted to say, however, that it's good *that* he's doing it and that this contains a certain element of gratitude. I suspect the more fundamental thought here would be that it's good that *someone* is doing it, and as a result it's good that he's doing it, and what we're grateful for that set of circumstances.
[69] He may, however, be concerned about what makes that a duty *for him*. That is, he can be unconcerned about the right-making features of helping schoolchildren but think that his years of doing so has placed a special burden on him to continue.

what he takes to be his duty will not be enough for that action to have positive moral worth and thus he won't be praiseworthy. Regardless of his concern for morality in general, or even his moral obligations in particular, he's not acting from the reason which makes it a good action or might make it his obligation. Even if it were a good reason to act, it wouldn't be *the* right reason for action on these accounts. The question, then, is whether this is correct diagnosis of this particular case. This is likely to be controversial since acting from the motive of duty is a prominent account when it comes to moral worth. Accordingly, anyone sympathetic to such an account might have clear intuitions that this is a good enough, or even *the* right reason for action. Those who think that a motive of duty is sufficient will judge the retiree praiseworthy and this will conflict with the CRT and PRMR. Even those who think that acting from the motive of duty *sometimes* justifies praise might think that this is one of those cases. One could think that acting solely from duty is praiseworthy in certain contexts, like when the obligation is strong, it's difficult to meet, and no one else is in a position to fulfill the obligation.[70] We can imagine, then, that the retiree acquired a strong obligation from years of being depended on by the community, that his mobility is weakened significantly at this stage of his life, and that no one else is available to fulfill the role of crossing guard.[71] If he's still determined to do his duty, even if only for the sake of duty, then his action and reasons may appear enough for praiseworthiness, at least in this context. There's motivation, however, to reject the idea that the retiree is

[70] Alternatively, one who accepts a view with these conditions (or others like) it might say that an adequately specified, or modified, version of a "motive of duty" account is sufficient for praiseworthiness.
[71] Although I won't discuss it here, it's also worth considering a variant of this case where the retiree *does not* have a strong obligation to remain in the role of crossing guard. One could think that such commitment to a weak obligation is an even better candidate for praiseworthiness than when there's a strong obligation.

praiseworthy because he acts from the motive of duty.

As we saw in section 4, acting from the "motive of duty" seems neither necessary nor sufficient for praiseworthiness. This allows us to say that the retiree doing the right thing because it's his duty is not itself enough for praiseworthiness. Filling out the case in such a way that it's both his strong obligation and that it's difficult to meet may seem to make him more admirable, though there's reasons to think that even this won't result in his being praiseworthy. It may be that acting solely from the motive of duty is not only the wrong motivation but it's also *objectionable*. Markovits, drawing in part from Michael Smith (1994) and Bernard Williams (1981), points out that "The Kantian 'truly moral man' seems guilty of a kind of moral fetishism… or at best, of having "one thought too many" (Markovits, 2010, p.204). Even filling out the case, the retiree seems to be focused on moral obligation in such a way that it's *getting in the way* of attending to the more direct reason that the action is good—say, keeping the children safe. We might also think that there's something off-putting about a person who, when considering *why* he ought to perform an action, thinks first or only of moral duty rather than the welfare of the children depending on him.[72] There are strong considerations then, for the claim that the retiree is not intuitively praiseworthy. RTRR accounts capture this and it appears that those sympathetic to a "motive of duty" account would need to either defend its sufficiency in the face of counterexamples, make the primary motivation of duty more palatable, or both.

[72] For more on this particular claim see, e.g., Wolf on Williams' famous "one thought too many" passage (Wolf, 2012).

3.7 Doing the Suboptimal for the Right Reason

The last case that I'll consider involves an agent performing a good action for the right reason when there exists some *better* action the agent could have performed. That is, cases I've labelled as *doing the suboptimal for the right reason*.[73] To illustrate this sort of case we can imagine a wealthy CEO, who has more money than she could ever spend and more than her descendants could ever need. While she cares about many charitable causes, and donates money each year, her motivation to help those in need could be stronger. As it is, she's disposed only to donate just a little more than she could write off as a tax deduction. We can assume that, given her circumstances and the causes she supports, the *right* thing to do would be to give much more each year. Still, she makes the donations because of the good that the money will do in the hands of the charitable organizations.

What makes this case interesting is that the CEO performs an action which is clearly good, and she does so for the right-making reason, it's just not the *best* course of action. It might seem overly stringent, then, if an RTRR account entails that we ought to withhold praise. To make this thought even stronger we could assume that it's *her* best action given her psychological makeup. She may be self-interested to the point that this sets limits on how much she's capable of attending to the needs of others, though not so self-interested that it prevents her from doing quite a bit of good. Is it appropriate to withhold praise even though we can't reasonably expect anything more than the good

[73] While, in a certain sense, any action that is not the best action would be "suboptimal," I'll be concerned only with suboptimal actions which are also good.

action she's already performed? My own intuition is that the CEO does deserve praise. One thing that would make this even more plausible is assuming that she's not at fault for having those competing self-interested desires. She may have grown up in poverty and so the instinct to lean towards preserving resources for oneself has become ingrained. It could also be that, although she recognizes that she currently has more money than she'll ever need, her background makes her overly (but understandably) cautious about finances. Withholding praise from the CEO in this case, then, appears too demanding.

This diagnoses initially appears to be inconsistent with RTRR accounts. While one's motivating reasons could coincide with the reasons morally justifying an action, these accounts also seem to require it to be *the* right action in order to be praiseworthy. Moreover, as we saw in section 4, the stipulation that one could be praiseworthy for performing "*a* good action" for good reasons isn't meant to relax the conditions for praiseworthiness. Rather, it accounts for cases in which there are equally good actions or equally good reasons. Although we haven't assigned values to the amount of money the CEO gives and the amount she could have given, this can be done in a way where it's clearly implausible that the two actions are equally good. One option here might be to relax the standard, at least when it comes to the kinds of actions for which a person can be praiseworthy. That is, holding that an agent could be praiseworthy if they *do a good thing for the right reason*. This would capture cases of *doing the suboptimal for the right reason* like the CEO, however it would seem to capture too many cases. Suppose, for example, that instead of donating a large sum of money to charitable organizations our wealthy CEO reaches out to a needy family in the community. As it

happens, she's the CEO of the cell-phone service provider that the family already uses and when she discovers this she wants deeply to help them. As a result, she reaches out to give them 50% off of their phone bill for the current month. We can suppose as well that she did this solely out of genuine concern for the well-being of the family. It seems here that the CEO does a good thing for the right reason but doesn't deserve praise. Although she may have helped, and was motivated by the right reasons, it just doesn't feel like she's done *enough*. Relaxing the standards for right action, then, won't be a solution. However, there may be another solution, or perhaps a compromise, at least for the CRT.

While Markovits' CRT doesn't itself build in degrees of praiseworthiness, she does make room for a scalar understanding of moral worth. In particular, she defends the idea that "right actions have moral worth to the degree that the noninstrumental motivations for their performance coincide with noninstrumental moral justifications for their performance" (Markovits, 2010, p.238). This is meant to hold, not only for best possible action, but also for suboptimal actions.[74] The emphasis would not be, as I've been characterizing it, equally placed on both the right reasons *and* the right action. Rather, moral worth would be determined by the reasons (both motivating and justifying) and could be had so long as the action is good (or, at least, that the agent believes the action to be good). Given that the CEO has performed a right action then, she could be said to be praiseworthy to the extent that her motivations matched the right-making reasons for the action. Now, on this view it may initially seem that she's just as praiseworthy as she

[74] And, in fact, this can hold even for *wrong* actions on Markovits' view (Markovits, 2010, pp.240-1).

would have been had she donated much more money. After all, what we're varying here is the kind of action performed and not necessarily her motivating reasons. We could even stipulate that her motivating reasons coincide with the justifying reasons to the same extent that they would have had she donated more. This would be a problem since, if there is a difference in praiseworthiness that depends only on the nature of the action, a scalar account of moral worth that only has implications from reasons for actions won't explain this. I think that the response to this worry should be that, even though the account is fundamentally about reasons, reasons and actions are connected in such a way that there *are* implications for actions themselves. In particular, that the CEO could not have given more money without her motivating reasons better coinciding with the justifying reasons. Her cautious self-interest was in tension with her motivation to do the right thing and so we might think that she *couldn't have* given more without shifting the weight of those motivations. After all, if her only motivation were the right-making reasons there seems to be no explanation for why she wouldn't have given all she could with the other features of the case in place. The CRT understood in this fuller way, then, seems to have an explanation for *doing the suboptimal for the right reasons*—that agents can be praiseworthy in such cases, but to a lesser degree than they may have been had they performed the optimal action.[75]

Whether or not Arpaly's PRMR can explain intuitions about the case in a similar manner is a bit trickier. The potential issue would proceed in much the same way, that

[75] A consequence of this may be that the CEO who only gives a small discount to the needy family also deserves praise. While I find this counterintuitive, this could perhaps be mitigated by the consolation that she deserves *very, very little* praise.

what we're varying here is the nature of the action and not the degree of moral concern. We would need to consider, then, the stipulation that the CEO is morally concerned to the same extent that she would be had she donated more money. Unlike before though, it may not be that varying the action would necessitate a change in what determines the degree of praiseworthiness. Having moral concern could be independent of the performance of actions in a way that motivation is not. Arpaly doesn't give a full account of moral concern but says that it's associated with motivational strength, emotional investment, and being "morally conscious" (Arpaly, 2002, Ch.3 pp.20-2). These latter two characterizations likely won't help when it comes to the CEO, however "motivational strength" could function just as well as a coincidence of motivating and justifying reasons. If a full account of moral concern allows for motivational strength to do the work in this case, despite no variation in emotional investment or extent to which the CEO is morally conscious, then PRMR may explain intuitions about the case just as well.

3.8 RTRR Accounts and Praiseworthiness

Accounts that tie praiseworthiness to doing the right thing for the right reasons do very well at capturing intuitions about cases. They provide natural explanations for straightforward cases like doing good with the best of intentions and doing what's right with bad motivations. They also plausibly explain *right outcome, wrong reason* cases like the selfish couple who adopts a child and the man who saves a stranger in the lake. The same can be said for cases of *refraining from wrongdoing with indifference*, like the young women who refrains from cheating on her exam. When it comes to even more

difficult cases these accounts also seem to have the resources to provide the intuitively correct diagnosis. However, some may have more resources, or provide simpler explanations than others when it comes to these further cases.

I've mostly set aside the RTRR accounts from the moral responsibility literature in order to focus on the more detailed accounts we find in the moral worth literature. One thing to note, though, is that there is one case in particular in which Wolf and Nelkin's accounts may have a certain advantage over the CRT and PRMR. Which conclusions we should draw about the case of the retiree who acts from the motive of duty turned on whether doing the right thing from the motive of duty was sufficient, or at least sometimes enough for praiseworthiness. Although I side with Markovits and Arpaly on this matter, those who are still sympathetic to both that Kantian view and RTRR accounts may have reason to prefer Wolf or Nelkin's account. Since they don't build in that the right reason is the reason for which the action is good, it leaves open the possibility that acting from duty is the (or at least one of) the right reasons.[76] Of course, the benefits of ambiguity here may be far outweighed by the benefits of a more precise account.

When it comes to *doing the suboptimal for the right reasons* it seems that Markovits' CRT account has advantages over Arpaly's PRMR. First, the way in which Markovits accommodates degrees of praiseworthy more obviously accounts for why the wealthy CEO is less praiseworthy than she might have been had she performed the optimal

[76] This is not to say that Wolf or Nelkin can't rule this out. In expanding on what "the right reason" involves (or, in Wolf's case, perhaps also "the True and the Good") the accounts could turn out much like that of Markovits or Arpaly.

action. While the PRMR includes the feature that the degree of praiseworthiness varies with the degree of moral concern, it would take spelling that out in more detail in order to apply it to the case of the CEO. This can likely be accomplished, though Arpaly would want to be careful that it doesn't merely collapse into Markovits' "degree of coincidence" account. Second, even if it can be done there still seems to be something desirable about the manner in which Markovits captures degrees of praiseworthiness. Namely, that the same fundamental feature is playing the key role both when it comes to determining moral worth and determining degrees of moral worth. With PRMR, on the other hand, there's an *additional*, partially independent aspect that determines the extent to which a person in praiseworthy.[77] On Markovits' view, degrees of praiseworthiness vary in direct relation to the coincidence of the motivational and justifying reasons, which is the fundamental idea of the CRT. We might think that this is kind of elegance is a theoretical virtue, and one that PRMR doesn't fully achieve.

There is, however, a place where we might think that Arpaly's account, and specifically the importance of moral concern, is helping to explain intuitions. Concerning the case of the young woman who refrains from cheating on her exam, both the CRT and PRMR agree that she's not praiseworthy. This accords with intuitions and so both accounts do equally well at diagnosing the case. Still, Arpaly's PRMR may be able to say *more* about the case than Markovits' CRT can. In discussing this case I suggested (though ultimately rejected) that temptation may be playing the key role in

[77] I use the modifier "partially" since Arpaly believes that motivational strength is an element of moral concern, and thus that component could connect up well with the more fundamental elements of the account.

intuitions about whether the young woman deserves praise. However, it may be that in cases of *refraining from wrongdoing with indifference* temptation is playing a more indirect role. That is, without temptation agents simply don't have the opportunity to demonstrate moral concern. If they're indifferent towards reasons for acting wrongly, and thus not tempted at all, then there won't be any aspect of morality to be concerned about. As a result, they won't merit praise. When agents *aren't* indifferent to the reasons for acting wrongly, often when there's temptation, their moral concern comes back into the picture and they can potentially be praiseworthy. Temptation may not be necessary or sufficient for praiseworthiness in general, but in certain kinds of cases it may come out that resisting temptation is an important feature. In particular, that in some cases it's what allows us to determine an agent's level of moral concern. If this is right, and Markovits can't capture this adequately without appeal to moral concern, then we might think that the "moral concern" clause of PRMR is doing important further explanatory work. Whether this would be enough to prefer PRMR over the CRT, or if we should prefer the CRT given its own advantages, is unclear. What we can say, however, is that each account has the resources to accommodate ordinary intuitions about difficult cases or, at least, make any counterintuitive judgments more palatable.

References

Arpaly, Nomy. (2002). *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford: Oxford University Press.

Foot, Philippa. (2002). *Virtues and Vices: And Other Essays in Moral Philosophy*. Oxford: Oxford University Press.

Markovits, Julia. (2010). Acting for the Right Reasons. *The Philosophical Review, 119*(2), 201-242.

Markovits, Julia. (2014). *Moral Reason*. Oxford: Oxford University Press.

Nelkin, Dana K. (2008). Responsibility and Rational Abilities: Defending an Asymmetrical View. *Pacific Philosophical Quarterly, 89*(4), 497–515.

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

Nelkin, D. K. (2016). Difficulty and Degrees of Moral Praiseworthiness and Blameworthiness. *Noûs, 50*(2), 356–378.

Smith, Michael. (1994). *The Moral Problem*. Oxford, UK: Blackwell Publishing.

Stratton-Lake, Philip (2000). *Kant, Duty, and Moral Worth*. London, UK: Routledge.

Strawson, Peter. (1962). "Freedom and Resentment." *Proceedings of the British Academy* 48: 187–211.

Williams, Bernard. (1981). "Persons, Character, and Morality." In *Moral Luck*, edited by Rachels, James, 1–19. Cambridge: Cambridge University Press

Wolf, Susan. (1980). Asymmetrical Freedom. *The Journal of Philosophy, 77*(3), 151–166.

Wolf, Susan. (2012). One Thought Too Many: Love, Morality, and the Ordering of Commitment. In *Luck, Value, and Commitment: Themes From the Ethics of Bernard Williams*. Edited by Heuer, Ulrike, and Lang, Gerald. Oxford: Oxford University Press