UNIVERSITY OF CALIFORNIA, SAN DIEGO

Responsibility, Reasons-responsiveness, and History

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

in

Philosophy

by

Craig Kushel Agule

Committee in charge:

Professor David Brink, Co-Chair
Professor Dana Kay Nelkin, Co-Chair
Professor Saba Bazargan-Forward
Professor Cathy Gere
Professor Samuel Rickless

2017

The Dissertation of Craig Kushel Agule is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____
Co-Chair

_____
Co-Chair

University of California, San Diego

2017

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I have many people to thank for assistance and support. I have learned much from the philosophy faculty at UCSD, and I have benefitted immensely from my many conversations with the faculty inside and outside of the seminar room. I want to offer special thanks to the members of my committee. At different times and in different ways, Saba Bazargan-Forward, David Brink, Cathy Gere, Dana Kay Nelkin, and Sam Rickless have given me invaluable feedback and support. I am particularly indebted to David and Dana, my co-chairs. Many of the best ideas in this dissertation stem from conversations with one or both of them, in a seminar they co-taught on moral responsibility, in independent studies, in feedback on various drafts, and elsewhere. They have been incredibly patient, generous, and helpful, and I thank them tremendously for that.

I also want to thank Professor Gary Herbert of Loyola-New Orleans and Professor Mark Anderson of Belmont University. Both of them were spectacularly welcoming to me as I began my path into philosophy as an outsider. My conversations with Professor Herbert about Kant's moral philosophy in the Loyola faculty dining room and my conversations with Professor Anderson about Plato's dialogues in Nashville will always remain some of my favorites, critical for helping me see the rigor, pleasure, and importance of philosophy.

I owe an immense debt of gratitude to my peers. I owe thanks to too many current and former graduate students at UCSD and elsewhere to list, but I want to mention a few for being particularly giving of time and attention: Matt Braich, Eric

Campbell, Kathleen Connelly, Cory Davia, John Dougherty, Stephen Galoob, Joyce Havstad, Gil Hersch, Alex Marcellesi, Noel Martin, Per Milam, Theron Pummer, Erick Ramirez, and Ben Sheredos. Their advice and feedback have made me a better philosopher and the arguments in here better philosophy. Of my peers, I save my greatest thanks for Amy Berg. From our visit weekend to today, she has been a steady source of intellectual and moral inspiration.

Finally, I want to thank the people in my life who have simultaneously supported and abided my becoming a professional philosopher. I want to thank Laura Everitt--I never would have taken the possibility of pursuing philosophy seriously without her support. I want to thank my siblings Rebecca, Jonathan, and Veronica. Every older sibling needs younger siblings to pressure him to be a fine exemplar (and to push back against sporadic hints of arrogance). I want to thank my father Richard for his enthusiasm and conversations, especially about grim cases of drunk driving. I want to thank Steve Wetzel, who served as an enthusiastic outside reader and sounding board; his willingness to work through the material and offer his thoughts was greatly appreciated. And finally, I want to thank my mother Jane. A son owes an endless debt to a mother. Almost all of the significant progress made on this dissertation was made during my trips home--and that reflects her goodness. Thanks so much.

Chapter 4 has been adapted from Craig Agule, "Resisting Tracing's Siren Song," Journal of Ethics & Social Philosophy 10(1):1-24 (2016). The dissertation author was the sole investigator and author of this paper.

Chapter 5, in part, is currently being prepared for submission for publication as Craig Agule, "Being Sympathetic to Bad-History Wrongdoers." The dissertation author was the sole investigator and author of this paper.

VITA

1999	Bachelor of Arts, University of Virginia
2002	Juris Doctorate, University of Virginia
2002-2004	The Staff Attorneys' Office of United States Court of Appeals, Fifth Circuit
2004-2009	King & Ballow Law Offices
2008	Master of Arts, University College London
2017	Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

Craig Agule and Carwyn Hooper, "Tobacco Regulation: Autonomy Up in Smoke?," Journal of Medical Ethics 35(6):365-368 (2009)

Craig Agule, "Resisting Tracing's Siren Song," Journal of Ethics & Social Philosophy 10(1):1-24 (2016)

ABSTRACT OF THE DISSERTATION

Responsibility, Reasons-responsiveness, and History

by

Craig Kushel Agule

Doctor of Philosophy in Philosophy

University of California, San Diego, 2017

Professor David Brink, Co-Chair
Professor Dana Kay Nelkin, Co-Chair

I argue for an ahistorical account of moral responsibility. An agent is responsible for her behavior, and is thus a fitting target of reactive attitudes such as praise and blame, if and only if the agent acts while in the possession of the reasons-responsiveness capacities to recognize reasons and to act in accord with the reasons she recognizes. We need not augment this account with conditions tied to the agent's history, as the account can provide satisfying explanations for cases of culpable incapacity (where the agent is incapacitated because of her prior

misbehavior) and of bad history (where the agent's wrongdoing is the product of having been neglected or mistreated). This ahistorical account of moral responsibility shows us that responsibility is about the sort of agent you are when you act, not about what happened in your past.

Introduction

Imagine two killers. The first killer, delirious after drinking and taking barbiturates for days, armed the whole time, shoots and kills a close friend. This killer has only a tenuous grasp upon the world and later has little recollection of the killing. The second killer had a horrendous childhood. He was abused and neglected by both his parents and civil institutions. As an adult, he shoots and kills two innocent teenagers in the course of a complex crime. This second killer is cold, calculating, and malicious, and he later considers compounding the wrong by taunting the victims' families.

Blame might seem called for in both cases. Intentional murder is deeply wrong, and these murders particularly so. In both cases, blaming these killers might be an important way that we could experience our moral repugnance at their behavior and recognize the moral importance of their victims. But it is important not to blame indiscriminately. Doing so might rob blame of its import, and blaming those who are not due blame seems intuitively wrong. It is important to get our blaming right.

Part of the story of getting blame right has to do with the nature of the behavior at issue, whether the behavior was wrongdoing or not. Were we to find out that one of the killings was somehow justified, perhaps in a complicated case of self-defense, that might make blame inappropriate. But another part of the story of the appropriate reactions has to do with the agent's relationship to the behavior. This is the question of responsibility. One powerful explanation of the conditions of

appropriate blame identifies responsibility with the possession of certain normative capacities, in particular the capacity to perceive moral reasons and the capacity to act in accord with one's perceptions. This is the reasons-responsiveness account of moral responsibility. The reasons-responsiveness capacity can explain why ordinary healthy adults are responsible but children and the mentally ill are not. The adults possess the normative capacities, but the capacities are still developing in children, and the capacities are compromised in at least some cases of mental illness.

On the reasons-responsiveness account, it might seem likely that the first killer is going to be excused but the second killer will be responsible and thus to blame. After all, at the time of their killings, the first killer was delirious, while the second killer was cold and calculating. These descriptions might lead us to suspect that the first killer's normative capacities were significantly compromised, but the second killer's were not. These apparent verdicts conflict with the reports many give in response to these two cases, especially when the cases are fully described. Many think that the first killer is plainly to blame and the second killer is at best partially to blame. This is because many think that any satisfying account of moral responsibility must take into consideration an agent's history in addition to whatever else it might consider.

So does history matter? Of course, an agent's history might explain why the agent is or is not reasons-responsive. Supposing that the first killer was not reasons-responsive, it stands to reason that his normative incapacity was the causal upshot

of his history of abusing alcohol and drugs. And an agent's history might be good evidence of the agent's current normative capacities. For instance, even if we did not have the report that the first killer was delirious, the history of alcohol and drug abuse might lead us to suspect that he was not fully reasons-responsive.

However, the core reasons-responsiveness account does not give history any direct role in the determination of moral responsibility, and so if the common responses to these cases are correct, these cases pose a problem for the reasons-responsiveness account of moral responsibility. Surely the first killer cannot point to his own reckless self-incapacitation to appeal for an excuse for murdering his friend, and yet the reasons-responsiveness account seems ill-prepared to explain why he could not make that appeal. The case of the first killer thus suggests that reasons-responsiveness is not necessary for moral responsibility, because it seems that the first killer is morally responsible for his killing despite not having been reasons-responsive at the time. Then consider the second killer. We might suppose it to be true that he was in control of his killings in the way articulated by the reasons-responsiveness account. But we might consider it no surprise that the second killer became a killer when we think of his childhood. We might think that anyone with those experiences might have become a vicious killer, and that realization seems to temper our blame toward the second killer. Thus, the case of the second killer suggests that reasons-responsiveness is not sufficient for moral responsibility, since his history tells against his moral responsibility despite his being apparently fully reasons-responsive.

Cases like these lead many theorists to insist that any adequate theory of moral responsibility must be historical, that is, that any adequate theory of moral responsibility must give history a fundamental role in the fixing of moral responsibility. Some historicists think that these cases mean that reasons-responsiveness should be abandoned for an alternative and historicist account of moral responsibility, while some historicists think that reasons-responsiveness can be modified to add a limited but important historicist element. I reject both of these moves. Instead, I defend an ahistorical, reasons-responsiveness theory of moral responsibility. An agent is morally responsible for her behavior if and only if she was reasons-responsive at the time. We do not have reason to give history any fundamental role to play.

My defense of this ahistorical view follows. In Part I of my dissertation, I defend the core reasons-responsiveness account. I first outline the sense of moral responsibility at issue: an agent is morally responsible for her behavior when the agent is a fitting target for reactive attitudes like resentment, indignation, and guilt. I then show that the reasons-responsiveness account provides a good explanation of when these attitudes are fitting: it provides an extensionally satisfying account for cases of widely accepted excuses like immaturity and insanity, and it provides a satisfying sense of the sort of control usually thought to be central to moral responsibility. In Part II, I turn to face the question of history. After outlining the sort of history at issue, I look at the two sorts of cases often thought to mandate a historical theory of moral responsibility. For both sorts of cases, I show that we can

satisfyingly explain our reactions to the cases and the agents' histories with the robust resources of the ahistorical reasons-responsiveness theory. This requires us to be careful about the behavior at issue, and it requires us to attend closely to the sorts of reactions at stake. But this careful, close examination shows that the reasons-responsiveness theory is itself satisfying, without any historicist element, even for these cases.

Showing that the ahistorical, reasons-responsiveness account can provide satisfying explanations of the cases often thought to compel historicism does not mean that historicism is false--some alternative defense of the fundamental relevance of history might be availing. However, these cases are often taken to be the strongest support for historicism, and so offering satisfying ahistoricist explanations of the cases shifts the burden of argument to the historicist, thereby giving the reasons-responsiveness theorist good reason to maintain ahistoricism.

Chapter 1 Reactive attitudes and responsibility

Praise and blame lie at the center of my concerns about responsibility and history. These are personally familiar attitudes to me. I have experienced being resentful towards someone for how they had behaved toward me, feeling the heat of my anger, my sense of the wrong, and the import I took their wrong to have for how I saw them, how I would interact with them. Although moral philosophers have long taken these attitudes to have special importance, they have been the subject of a renewed focus since P. F. Strawson's landmark 1962 essay "Freedom and Resentment."[1] For Strawson and the many philosophers who follow him, these attitudes are the attitudes we take by virtue of seeing each other's behavior as expressing value. As Strawson writes, these "participant reactive attitudes are essentially natural human reactions to the good or ill will or indifference of others towards us, as displayed in their attitudes and actions" (2008, pp. 10–11).

Accordingly, I begin my investigation by offering an account of the reactive attitudes and their relationship to my theory of moral responsibility. This account is in significant part a synthesis of the extant accounts which populate the moral responsibility literature. My reliance upon the account is supported by its popularity and by its intuitive fit with my own experiences of praise and blame (and hopefully likewise with my readers' experiences of praise and blame). This account is for the most part stipulative, though hopefully its usefulness in addressing the issues raised through the course of this dissertation provides it further support.

[1] I cite the version of this essay reprinted in Strawson (2008).

1.1 The elements of the reactive attitudes

In this section, I present my account of the reactive attitudes (and especially of praise and blame). I begin with three widely recognized characteristic elements of reactive attitudes: their propositional content, their phenomenology, and their associated behavioral dispositions. In addition to these elements, I follow several recent scholars in recognizing a fourth, perceptual element. The reactive attitudes frame our perception of the agents to whom we are reacting.

In developing these four features of the reactive attitudes, I take inspiration from the recent literature seeing the reactive attitudes as a particular, complex type of emotion. This literature--and especially the work of Lucy Allais (2008a, 2008b), Elisa Hurley and Colleen Macnamara (2010), Macnamara's later independent work (2015), and Leonhard Menges (2017)--has made clear the importance of the distinctively emotional aspects of the reactive attitudes, and I follow those philosophers in that emphasis.[2]

[2] While I believe that it is true that the reactive attitudes are complex emotions, I do not take that to be a premise from which I might establish the elements of the reactive attitudes. Rather, my claims regarding the elements of the reactive attitudes are largely taken from introspection and from support from within the moral-responsibility literature. The supposition that the reactive attitudes are complex emotions might buttress my conclusions regarding the elements of the reactive attitudes, but it does not provide the central support. Because of this, and because a significant part of my project is to pick out the concepts at issue in talk of praise and blame, I leave for later the work of checking the fit between philosophers' work on the reactive attitudes and psychologists' work on putatively similar phenomena. I thank Dana Kay Nelkin for pointing me to the importance of this separate project.

1.1.1 Manifested quality of will

The reactive attitudes have propositional content: they are sensitive to and expressive of our perceptions and assessments. As Strawson explains, we can see the propositional content in "the great extent to which our personal feelings and reactions depend upon, or involve, *our beliefs about [agents'] attitudes and intentions*" (2008, p. 5, emphasis mine). That the reactive attitudes are content-sensitive is particularly clear from the way that our attitudes dissolve or change when we correct our beliefs about others' attitudes and intentions. My blame weakens or dissolves when I learn that an apparent wrongdoer bore no ill will but rather had intended to do right.[3] And my praise dissolves when I learn that an apparent rightdoer was acting on the basis of self-serving attitudes. My praise and blame are beholden to my judgments about the facts, revealing that my praise and blame bear content.

The nature of the content at issue can vary. While Strawson takes quality of will to be the content, R. Jay Wallace (1994) explains that the reactive attitudes are our reactions to our assessment of how agents' behavior meets our normative expectations. In general, however, contemporary Strawsonians hold that the reactive attitudes regard quality of will, and so we can see praise and blame as reactions to good will, ill will, or indifference displayed in actions and attitudes. This means that the Strawsonians should provide an account of quality of will and an account of manifestation.

---

[3] It might weaken rather than dissolve in cases where I think that the agent's failure to live up to his good attitude was due to some other insufficient care or concern.

Though there are likely non-behavioral phenomena which can manifest quality of will, quality of will is manifested most commonly in action.[4] Eliding significant complications, we can understand many of our actions as being the non-accidental upshots of our quality of will. For instance, an agent who is cowardly might therefore be particularly disposed to notice risks and threats and likewise particularly disposed to be motivated to avoid risks and threats. When these dispositions result in the agent noticing and acting to avoid a particular risk or threat, it seems that the avoidance behavior manifests the agent's cowardice. By contrast, consider a contrived case where the cowardice plays a merely causal role in the agent's performing some role: the agent, in running from one risk, is accidentally led to some important find. The find is the product of the agent's cowardice, but it does not seem to manifest the agent's cowardice, not in the right sort of way. The Strawsonian account of the propositional content appeals to this intuitive distinction between the running and the find, where the latter is merely caused by quality of will, while the former manifests quality of will.

[4] Here, I elide the distinction between actions and omissions. For many Strawsonians, including myself, it seems true that we can often discern an agent's quality of will in both her actions and her omissions, and this makes her omissions grist for the reactive attitudes. In an example offered by Michael McKenna, Casper cancels his weekend business plans to enjoy time with his friends, but he does not consider whether he could use his free time to tend to his recently ill daughter (2012, p. 60). McKenna explains that, while desiring to spend time with friends is innocuous, we might nonetheless see Casper's choices as manifesting ill will because they reveal insufficient regard for his daughter. I accept as intuitively plausible, without hereby defending, that we might treat inaction in this way as similar to action.

That an action manifests the agent's quality of will is one way to understand the attributability sense of responsibility identified by Gary Watson (1996). Responding to Susan Wolf (1990), Watson distinguished between responsibility as attributability and responsibility as accountability. I'll return to responsibility as accountability later; for now, focus on responsibility as attributability. This is the sort of responsibility required for aretaic judgments. An aretaic judgment about an agent is a judgment about quality of an agent's values or character. For example, in saying that an agent's behavior shows her to be a coward, we hold the agent responsible for the behavior in the attributability sense. But for this to be the case requires more than that the behavior is the mere causal upshot of the agent's character or quality of will. As Watson writes, "The significant relation between behavior and the 'real self' is not (just) causal but *executive* and *expressive*" (1996, p. 233, emphasis in original, footnote omitted). Watson then argues that it is this executive and expressive nature of human (agential) behavior which distinguishes us from objects and from other creatures; likewise, I claim that it is these executive and expressive features which distinguish our manifesting behavior from our accidental behavior.[5] In order to support the cognitive assessment required for the first element of the reactive attitudes, we must take the agent to be responsible in

[5] In further developing the account of manifestation, we might appeal to the concurrence literature in the criminal law. The overwhelming majority of crimes are defined by a combination of mental and behavioral elements. Merely satisfying those elements, however, is not sufficient. The behavioral elements must be related in the right way to the mental elements, largely to avoid cases of accidental causation. For a good overview of the concurrence requirement with a nice argument for seeing the concurrence requirement in a normative fashion as opposed to a merely counterfactual fashion, see Alex Sarch (2015).

the attributability sense. Seeing the agent as responsible in the attributability sense, then, is necessary for the reactive attitudes.

The Strawsonian thus owes an account of quality of will. On a broad read of quality of will, quality of will might be our regard for anything at all which matters. Michael McKenna (2012) offers a broad account along these lines. For him, the quality of will manifested in action is the regard manifested in the behavior for other agents, for the agent herself, and for other relevant moral considerations. On this account, we might blame another agent for his faulty regard for important works of art or important features of the environment even if that faulty regard does not adversely affect anyone else and even if that faulty regard does not extend to any faulty regard of any other agent. Alternatively, we might take on the interpersonal commitments Strawson focused on and limit quality of will to regard for agents. As Strawson wrote, we care about whether "actions … reflect attitudes *towards us* of goodwill, affection, or esteem on the one hand or contempt, indifference, or malevolence on the other" (2008, pp. 5–6, emphasis added). And we could seek a unifying explanation. For instance, even if regard for agents is all that matters for the quality of will relevant to the reactive attitudes, other things, such as artwork or the environment, can matter derivatively where they matter to some agent. Thus, for here, I take on the Strawsonian, agent-focused account, leaving for later full consideration of this debate.

The reactive-attitudes theorist should also have something to say about the nature of regard. In one sense, our regard for things (reading things broadly) is our

set of dispositions to act in appreciation or response to such things. As George Vuoso explains:

> A person's character in this sense may be described as the collection of many of his dispositions to act. It is a subpart, though a very significant subpart, of one's personality. It encompasses or involves certain traits, such as honesty, loyalty, kindness, fairness, ruthlessness, and greed, but not others, such as intelligence, sophistication, nervousness, and sense of humor. For the purposes of moral evaluation, it is the measure of a person. (Vuoso, 1987, p. 1670)

We can see here the affinity with Watson's attributability sense of moral responsibility. Thinking of quality of will merely in terms of behavioral dispositions is not ultimately satisfying, however, without an explanation of those dispositions. Honesty is not merely a statistical disposition to tell the truth; it is a disposition to tell the truth rooted in something more significant about the agent. Accordingly, we need something like David Shoemaker's (2015) fuller account of the sorts of things which might make up this regard. For Shoemaker, quality of will is comprised of three different classes of phenomena: clusters of cares and commitments, evaluative judgments, and sensitivity to interpersonal value. The honest person is disposed to act honestly not accidentally but rather because the honest person cares about the values in truth-telling, finds those values relevant and important, and is sensitive to the right to frankness held by others.[6]

[6] Have the dispositions dropped out of the picture? Why not take quality of will simply to be the cares, concerns, and the like picked out by Shoemaker? I do not see any reason to deny that identification, though it is worth remembering that the resulting behavioral dispositions might be an important part of the explanation of why a Strawsonian might care about others' cares, concerns, and the like.

There are two related puzzles suggested by this account of the sort of quality of will involved in reactive attitudes like praise or blame. We ordinarily focus on but some aspect on an agent's quality of will in blaming or praising them; we rarely praise or blame an agent for being a good or bad agent overall. Instead, we praise them for this good action or blame them for this bad action. So in that sense, we do not praise or blame the agent for their quality of will in its totality. Relatedly, we can praise or blame agents for behavior which is out-of-character. We might, for instance, blame an otherwise-honest agent for an out-of-character lie. In that case, it might seem that we are blaming the agent despite, and not on account of, their quality of will.

We must keep two related features of the Strawsonian account in mind to address these concerns. First, we praise or blame agents for the quality of will manifested in some particular action, and that quality of will need not be (and indeed rarely will be) the entirety of their quality of will. Second, the Strawsonian appeal to quality of will does not require a commitment to a unified and consistent psychology. The will of any real agent is almost certainly diverse and complicated, inconsistent in important ways. That means that an agent's broad quality of will might be generally honest, but in some particular case, that general honesty is not the quality manifested in an instance of lying. These two features can explain why mostly good agents sometimes act poorly, even in the matters which most mark them as good agents, and why poor agents sometimes act well. Thus, when the

Strawsonian talks of the quality of will at issue in the reactive attitudes, we must be careful to pick out the quality of will at issue in the particular action.

There are ordinarily two agents involved in the reactive attitudes, the agent who acts and the agent who reacts.[7] The propositional content of the reactive attitudes tracks the quality of will of the agent who acts. But we can then ask about the psychology of the agent who reacts. That the reactive attitudes have propositional content does not require that the agent experiencing the attitudes have either conscious recognition or a full-throated judgment of that content. It seems familiar that we often blame without reflection. It seems that I sometimes even blame without realizing it--only coming to see that I've been blaming when my attention is brought to my own behavior after the fact. This suggests that the reactive attitudes do not require conscious, deliberated judgment. David Zimmerman (2001) gives us a broad range of the sort of cognitive options which might suffice: explicit propositional beliefs, quasi-intentional states, vague beliefs with only roughly specified truth conditions, and patterns of intentional salience (which may or may not be readily formulable as beliefs). In all of these cases, there is some phenomenon internal to the reacting agent that we might identify as a reflection or recognition of the target agent's quality of will.

Allowing that phenomena less than fully fledged judgments can suffice for the propositional content of the reactive attitudes allows us to easily sidestep the

---

[7] We can praise and blame ourselves. In that case, there is one agent, who is both acting and reacting. Still, we can separate our examinations of that agent into the two roles.

wayward reactions problem. Is it possible for me to blame someone if, when I reflect honestly, I judge that their behavior involves no wrongdoing? Wallace (1994, pp. 42–45) asks us to imagine someone who was raised to think that premarital sex is improper but who has since abandoned that belief. When the agent becomes aware of someone who has had premarital sex, they may experience something like resentment. However, if resentment, as a reactive attitude, requires judgment of a robust nature, and if we trust the agent with respect to her own judgments, then this cannot be resentment. If we trust the agent that she judges that premarital sex is morally permissible, and if we accept that an agent cannot be marked by conflicting judgments, we must conclude that the agent cannot also judge that premarital sex is morally impermissible, and this conclusion means that the reaction the agent is experiencing cannot be resentment, lacking the necessary propositional content.

We might allow that cases like this pick out wayward resentment, and we might then conclude that resentment and wayward resentment are two different phenomena. In cases like this, we'd have wayward resentment, which is like real resentment, but propositionally defective. But what is to be gained from this ontological expansion? Or perhaps we might abandon the idea that the reactive attitudes' propositional content has a cognitive, occurrent element. If there is no cognitive element, then wayward resentment and ordinary resentment might both be fully resentment. Shoemaker (2015, p. 89) suggests this, focusing on anger rather than resentment in large part because of the cognitive implications of resentment. Wallace's response to cases of wayward resentment is particularly creative: the

reactive attitudes are associated with normative expectations, and normative expectations do not require judgments. His solution, however, leads him to define normative expectations in terms of reactive attitudes. Although the solution is not false merely because it is circular, the tight nature of the circle makes the solution unilluminating.

The better solution is McKenna's.[8] Once we allow that the cognitive content of the reactive attitudes need not be the sort of high-level judgments for which we feel immediate pressure of consistency, we can explain how we could both judge some action to be not improper and at the same time resent someone for the action. As McKenna writes, it isn't that these cases of wayward resentment are not true resentment: "It's rather that in such cases, they are linked to beliefs not fully endorsed or in some way defeated by way of the agent's wider class of beliefs, including her evaluative commitments" (2012, p. 67). Understanding the wayward resentment cases in this fashion reinforces the assessment that the propositional content of the reactive attitudes can take the wide range of forms pointed to by Zimmerman.[9]

---

[8] In email correspondence, David Brink has suggested a similar solution to this problem, writing that we might understand conflicts between resentments and contrary judgments as reflecting "a conflict of beliefs that reflects the operation of conscious processes and automatic processes, including bias." Both McKenna and Brink solve the potential problem by rejecting the idea that an agent cannot be marked by simultaneous, conflicting beliefs.

[9] A similar problem-and-solution pair arises for the emotions more generally. Consider fear. We might persist in our experience of fear even if we are, in some significant sense, assured of our safety. Think of the experience of fear during a scary movie or during a thrilling theme park rollercoaster. We might seem to face an unfortunate trilemma: is it that the agent in the theater judges herself to be in

To sum up: the reactive attitudes have propositional content, the quality of will manifested in the behavior reacted to. This quality of will is a matter of regard for things which matter, though exactly which things matter is for later exploration. We should locate this propositional content inside the agent experiencing the reactive attitude, though this does not require consciousness, awareness, or global consistency with other beliefs and attitudes. It requires something cognitive, but not particularly much.

1.1.2 The characteristic experiences

Much of the scholarly attention on reactive attitudes focuses on the propositional content of the attitudes and its psychological instantiation. But, as Manuel Vargas notes: "If we focus on just the cognitive aspect of blaming, we run the risk of presenting a pallid picture of responsibility, drained of its characteristic affect" (2013, p. 119). Thus, in addition to content of the reactive attitudes, we should recognize that the reactive attitudes also have characteristic phenomenologies. The distinctive phenomenologies of the reactive attitudes are particularly noticeable with the negative reactive attitudes. Resentment, for example, ordinarily feels agitated, tense, and metaphorically hot, and guilt feels metaphorically heavy. And, as Menges (2017) notes, these first-person experiences

---

significant danger (which would require us to think that the agent who willingly goes to the movie is in some significant way irrational), is it that the experience that results is wayward fear, or is it that fear does not have the significant connection to the perception of danger we might have expected? If we allow that our cognitive states can conflict, we might accept both that we have some experience of the perception of danger even as we at the very same time judge ourselves to be in no danger. That experience of the perception of danger can explain the persistence of fear as an emotion in the absence of the judgment of danger.

of the reactive attitudes track externally measurable phenomena like increases in heart rate and skin temperature.

While there are experiences characteristic of the various reactive attitudes, we need not assume either that there is some single characteristic experience for any particular reactive attitude nor that every instance of any reactive attitude must be associated with some distinctive experience. As Allais (2008a) explains, we should think of each of the reactive attitudes as being associated with a range of feelings, the particular feeling elicited at any time being a function of the reactive attitude at issue, the agent who is experiencing the reactive attitude, and the attendant circumstances. Anger might sometimes leave us agitated, but it might also sometimes leave us collected, perhaps depending upon whether anger is combined with the possibility of redress and improvement or not.

And no particular experience is required of any particular occasion of a reactive attitude. For instance, we might be resentful without any noticeable feeling. In such a case, we judge the other has wronged, we are disposed to act on our judgment, and we see the agent in light of that judgment, but there is none of the distinctive heat commonly associated with anger. Of course, it might be that there is a feeling in these cases, but it is a subtle one, readily overlooked. Perhaps there is both heated and cool anger, and cases that had seemed to lack distinctive phenomenologies are actually cases of cool anger. If so, we might persist in seeing the characteristic phenomenologies as necessary elements of the reactive attitudes. But another possibility is that the phenomenologies are ordinary but not necessary.

The three other elements are sufficient to pick the reactive attitudes out as a distinctive subject of interest.

1.1.3 The characteristic behavioral dispositions

The third commonly recognized element of the reactive attitudes is their associated behaviors. There is a "distinctive syndrome" (Wallace, 1994, p. 24) of practices characteristically associated with each of the reactive attitudes. When I resent, I might lash out at the target of my resentment, I might make clear to them how they have done wrong, I might seek to have them punished, or I might withdraw from social interaction with them. These are all behaviors familiarly associated with resentment. Here, as with the characteristic phenomenologies, the behaviors are characteristic and variable. As McKenna notes, while there are these characteristic practices, there is "no simple formula" (2012, p. 67). The behaviors involved in any particular case are likely to be highly variable, dependent upon the wrongdoing, the attendant circumstances, the relationship between the parties involved, and the like. Moreover, these are merely dispositions. Accordingly, no particular instance of resentment need involve any behavior at all. You can resent in private, and you can resent someone you do not and will not interact with.

1.1.4 Seeing-as

These three elements--propositional content, characteristic phenomenology, and characteristic behavioral dispositions--are widely taken to mark the reactive attitudes, though the exact details vary. I take on board this widely accepted account. However, these three elements alone cannot provide a sufficient sense of

the sort of praise and blame I am interested in. For my purposes, there is a fourth element of the reactive attitudes, one that seems often overlooked but also critically important: the reactive attitudes frame our perceptions, especially our perceptions relating to the target of the attitudes.[10] Because of the first element of the reactive attitudes, their propositional content, the reactive attitudes are widely seen as downstream from our perceptions; because of this fourth element, we should see that the reactive attitudes are also upstream from our perceptions.

How might it be that the reactive attitudes affect how we see the world? The reactive attitudes change what we take to be salient, and they change which interpretations we find compelling. When I react to another's wrongdoing (e.g., when I find that I judge that someone has done me wrong and concomitantly feel the characteristic heat of resentment), I am relatively more likely to notice other instances of their wrongdoing and I am relatively less likely to take seriously contrasting evidence of the other's good will. In this fashion, our reactive attitudes order the evidence we are presented about others' quality of will. As David Zimmerman explains:

> Your reactive emotions--resentment, anger, the temptation to forgiveness, the warmth of friendly nostalgia, and so on--play a distinctively *cognitive* role throughout in regulating your inquiry into the question of whether X has in fact been perfidious. How? As Rorty and de Sousa would put it, by *shifting your attention* from one set of facts to another, and then back over the same set of facts, with more care the second time around; by inducing you to see *patterns of salience* in the facts which you would otherwise have missed; by

[10] Allais (2008b), Hurley and Macnamara (2010), and David Zimmerman (2001) offer particularly helpful explorations of the way that the reactive attitudes' affective nature changes how we see the world.

> sometimes slowing you down and sometimes speeding you up, as genuine opportunities for inquiry ebb and flow. (2001, p. 532)

As Allais explains: reactive attitudes involve "seeing [another] in a certain way, being disposed to have characteristic patterns of attention, interpretation and expectation with respect to her actions" (2008a, p. 185). Thus I suspect that the framing effect of the reactive attitudes is in fact very broad. When I resent someone, I am therefore increasingly likely to notice flaws of all kinds and increasingly less likely to notice virtues of all kinds. For example, I am more likely to notice and be bothered by someone's grating voice when I resent them for some wrongdoing, and I am less likely to notice and be impressed by their sharp attire.

Taking seriously the framing element of the reactive attitudes can explain a number of features which seem to mark the attitudes. First, this element can help to explain the impenetrability of the reactive attitudes. So long as we are in the throes of the reactive attitudes, we are disinclined to notice competing evidence, and we are inclined to interpret the features we do notice in ways consistent with our reactions. This cannot completely explain why we can resent someone at the same time that we acknowledge that they've done no wrong, as in such a case we do notice the competing evidence and we do take it to have normative significance. But it can explain why the reactive attitudes are relatively resilient and self-reinforcing.

Second, the seeing-as element of the reactive attitudes explains why it is that we resent wrongdoers for their wrongdoing, rather than merely resenting them or merely resenting the wrongdoing. This is not to deny that we do not sometimes

assess either the agent herself or the action itself. We certainly assess those things.[11] And when we are assessing the agent, we might look to her behavior as evidence of her quality of will. For instance, we might use an agent's acting rudely once as evidence that the agent is herself rude. Likewise, we might use our impressions of the agent as evidence regarding the quality of a particular bit of behavior. For instance, we might use our judgment that an agent is generally rude to help us decide that a particular, ambiguous act was a rude act. However, there is a natural sense of resentment and many of the other reactive attitudes which involves both the agent and their behavior essentially, neither as merely evidence for the other. That there are essential roles for both the agent and the action can explain why resentment displays ordering effects. If, for instance, we reacted to others' quality of will and their behavior was merely evidence of that underlying phenomenon, then it is not clear that we should experience reactive attitudes so attuned to the particular behavior at issue, and it is not clear that the order of behaviors should make such a difference to the reaction to the ultimate bit of behavior.[12]

---

[11] As David Brink has suggested in correspondence, we might have good reason sometimes to separate these assessments. In parenting, for instance, it might be fitting to assess a behavior but to refrain from deep assessments of quality of will or agent.

[12] Consider a puzzle raised by David Hume (2000 bk. 2, pt. 2, sec. 2): how can it be right to justify blaming an agent now for some action in the distant past? Hume thinks that we are justified in this connection because (and presumably therefore when) the action stems from some persisting character trait, a trait that the agent had both then and now. But I'm no more satisfied with this explanation than Wallace (1994, pp. 122–23) is. We often blame people even if they've changed in the meantime, and we often blame people when their actions were out of character. Moreover, Hume's answer comes too close to seeing the behavior as mere evidence of the untoward character trait.

The seeing-as element readily explains this feature of the reactive attitudes. When we react to others on account of some bit of behavior, we see them in light of that bit of behavior. We are not seeing them in their entirety, judging them in the light of all of the evidence (indeed, that is plausibly something we never could do). Return to the ordering problem. Consider two otherwise identical agents, one of whom does some good behavior first and then some bad behavior second, and the other of whom does the same behaviors, but in the opposite order. The seeing-as element can explain why can we have such strikingly different reactions to the agents after their second behaviors. We see the first agent in light of their bad behavior, causing us to overlook their prior good behavior, and so we might strongly resent them, whereas we see the second agent in light of their good behavior, causing us to overlook their prior bad behavior, and so our resentment of them for that prior behavior might be mitigated or even absent. Thus, understanding the reactive attitudes as involving this seeing-as element seems to explain a common phenomenon of our reacting practice.[13]

[13] Dana Nelkin has pointed out an interesting puzzle raised by the combination of recognizing potential ordering effects and recognizing the seeing-as element of the reactive attitudes. If the seeing-as element of the reactive attitudes means that our perceptions of the later behavior are framed by our reactions to the earlier behavior, we might expect a different ordering effect. Consider the agent who behaves first poorly and then properly. Why not think that the resentment sparked by the poor behavior will lead us to underappreciate the later proper behavior? This is an interesting question. I find it plausible that resentment might sometimes work this way and sometimes work the way that I described in the body text here. Perhaps the particular ordering phenomenon witnessed depends upon which behavior is presented to us first, rather than which behavior occurs first. That is a different sort of ordering effect, though there too the seeing-as element can play an

Seeing the importance of the reactive attitudes as framing mechanisms is also confirmed by the recognition that the reactive attitudes are broader than just the negative reactive attitudes. The philosophical discussion has focused on the negative reactive attitudes.[14] Philosophers are fascinated by blame. We see in Wallace a substantive argument claiming there is a set of negative reactive attitudes worthy of distinctive attention. His account of the reactive attitudes is closely tied to the Strawsonian moral demand (and the concomitant obligation) that others regard us (and therefore treat us) with due good will, and so he limits the reactive attitudes to those connected with the belief that a moral obligation has been violated. He allows that we might react to the ways that others exceed those obligations, but "the reactive attitudes are explained exclusively by beliefs about the violation of moral obligations" (1994, p. 38). Wallace-style skepticism about the positive reactive attitudes is buttressed by the relatively thin phenomenologies and behavioral upshots of the putative positive reactive attitudes in many cases.

But this is too narrow. While the negative reactive attitudes might as a matter of contingent fact be particularly salient, "'positive' attitudes like gratitude can also be understood as responses to the way in which an agent demonstrates the character of her will in her actions" (Allais, 2008a, p. 184). Consider a case where

---

explanatory role. I will have to think further about the relationship between the seeing-as element and potential ordering effects.

[14] This is not to say that philosophers do not address praise significantly. Some philosophers, e.g., Wallace, limit their accounts to blame and resentment. Others, like Fischer and Ravizza, make broader arguments despite a strong focus on cases of blame. But some, e.g., Nelkin (2011), do devote significant attention to praise and to the comparative relationships between praise, blame, and moral responsibility.

someone has gone well beyond your expectations, perhaps a friend has put in an extraordinary effort to help you through a difficult spot, even at great cost to themselves. Their behavior expresses their tremendous regard and concern for you. Perhaps this leads you to feel a certain warmth and softness and to express gratitude and thanks to them. But even if not, or even if these reactions are subtle and difficult to notice, your reaction to their aid should color the way you see them. That is certainly my experience. While the aid is but one act out of an entire lifetime, if I attend to the act, I am inclined to notice other, consistent acts of good will, and I am inclined to interpret other behaviors in a sympathetic light. On my account of the reactive attitudes, if you do not experience these framing effects, you are not praising, at least not in the sense picked out by the reactive attitudes. But because we can make sense of the epistemic framing, we can make sense of the intuitively plausible positive reactive attitudes, despite their often behavioral and phenomenological thinness.

1.1.5 Episodes versus stances

Many of our ordinary experiences with the reactive attitudes can be explained by way of this account, but not all of them. Some of our reactive attitudes are brief, especially when we react to strangers for insignificant interactions. But many of our reactive attitudes are long-lasting. We might, for instance, resent someone close to us for a significant wrongdoing for quite some time. However, that does not mean that we need have their wrongdoing on our minds throughout or that our epistemic contact with the world is constantly mediated by their wrongdoing.

Instead, we should distinguish reactive-attitude episodes and reactive-attitude stances.[15] An episode is the immediate experience of an attitude, whereas a stance is the durable disposition to episodes of the attitude.

Thus, for instance, consider Adam who resents Beverly. While thinking of Beverly, Adam finds himself in a resentful episode. He is thinking of Beverly and her wrongdoing, he feels heated, he is disposed to punish Beverly or to lash out at her, and he sees the world colored by her wrongdoing. But on other occasions, when Adam is not thinking of Beverly, he does not feel heated, he does not experience any urge to punish Beverly or to lash out at her, and, importantly, his understanding of the world is not informed by thoughts of Beverly or her wrongdoing. This does not mean that he does not resent her. Rather, it means that he is not in an episode of resentment. He is, rather, in a resentment stance toward Beverly.

The exact relationship between episodes and stances is grist for more inquiry. For instance, might a reactive attitude stance involve more than a thin disposition to the corresponding episodes? Or: could an agent be said to be in the reactive attitude stance even if they never experience an episode? I set those questions aside. For my purposes, it is important only to recognize this intuitive distinction between two common ways to experience the reactive attitudes.

1.2 An important sort of blame

Although Strawson does not seem to have been particularly concerned with blame, the negative reactive attitudes pick out a distinctive and important sort of

[15] Here I borrow terminology from Menges (2017).

interpersonal blame. They have both the right backward-looking focus on wrongdoing and a sufficiently unwanted sting. That the negative reactive attitudes can be seen as an important sort of blame is part of the case for making the reactive attitudes central to the study of moral responsibility.

I do not mean here to take a particular stand on the essence of blame, on whether there is a single unified sense of blame, or on whether the negative reactive attitudes mark the most important sense of blame. There are many sorts of blame. There are behaviors which seem to be sorts of non-moralized blame. We blame the thunderstorm for the canceled picnic. While that indicates some sense that the thunderstorm has played a causal role in things being worse than they might otherwise be, we do not think that the thunderstorm has done anything wrong.

And there are several sorts of moralized blame which are not equivalent to the negative reactive attitudes. Strawson contrasted two ways of seeing others: akin to patients, whom we might diagnose and treat, and as morally responsible agents, whom we might praise and blame. When we see others as lacking the right sort of role in their own behavior, or when we see their character as the product of external and often malignant factors, we do not treat them as morally responsible for their behavior. Still, there is a sense in which we might be thought to blame agents we consider in an objective sense in cases where we judge them to do wrong, so long as we point to them and to their character in our explanations of their wrongdoing. And punishment might be another sort of moralized blame. While punishment might be thought to be inclusive or derivative of the reactive attitudes, it involves

the further necessary element of the intended attempt to impose harm, an element not needed in ordinary instances of interpersonal blame.[16]

The negative reactive attitudes are a distinctive sort of blame. Like the other sorts of blame, the negative reactive attitudes involve an explanatory assessment of some sort of negative phenomenon; here, it is the attribution of some bit of wrongdoing to poor quality of character.[17] And this is a backwards-looking assessment--indeed, the reactive attitudes are reactive. The negative reactive attitudes fit nicely in between merely cognitive blaming that involves only judgments and behavioral blaming such as requires punishment or change in social relationships. And the negative reactive attitudes seem widespread and familiar. Accordingly, we should see the negative reactive attitudes as constituting a distinctive and important sort of blame.[18]

Because the negative reactive attitudes do not necessarily involve any withdrawal of social interaction, punishment, or the like, it might be thought that

---

[16] Importantly, an advocate of a reactive-attitudes account of blame need not thereby also be a retributivist, thinking there is an immediate good in the imposition of some harm on a wrongdoer or the experience of some suffering by the wrongdoer. These views are consistent, but neither entails the other. For more on the relationship between a theory of responsibility and a theory of desert, see especially McKenna (2012) and Nelkin (2014).

[17] The relationship to the prior wrongdoing is an important sense in which the negative reactive attitudes are properly backward-looking, by contrast with the results-oriented concerns of consequentialist theories of appropriate blame like that offered most famously by J.J.C. Smart (1961) or the results-oriented concerns which motivate contemporary theories of quasi-blame like those offered by Derk Pereboom (2001, 2014) and Vargas (2005a, 2006, 2013).

[18] Skeptics not convinced by the argument which immediately follows can treat the bulk of this essay as an essay on the conditions of the reactive attitudes being appropriate as opposed to an essay on the conditions of a certain sort of blame being appropriate.

they lack the characteristic "sting" of blame.[19] This objection has been described forcefully by Pamela Hieronymi:

> it is unclear how the affective accompaniment of a judgment could, itself, carry the characteristic force of blame. An affective accompaniment of a judgment would be a certain unpleasant emotional disturbance, occasioned by the judgment. But, the force of blame seems deeper, more serious or weighty than simply being the object of certain unpleasant emotional disturbance. The affect, itself, seems insufficiently robust. (2004, p. 121)[20]

Consider punishment. Punishment's propositional content resembles resentment's, but punishment combines that content with the intentional infliction of pain or deprivation. That infliction seems sufficient for the characteristic sting of blame.[21] Likewise, we can see informal, social alienation as being marked by a sting--the sting of exclusion, degraded relationships, public excoriation, and the like. It is intuitive that blame "stings," and it might not be immediately clear how the reactive attitudes--especially given that they might be entirely private--sting. This can give us reason to doubt that the reactive attitudes constitute blame at all.

One way to locate a sting in the negative reactive attitudes is to see the reactive attitudes as important for our relationships of mutual regard. This is similar to Hieronymi's strategy; she locates the characteristic sting of the judgments of

---

[19] Because I think that the reactive attitudes do sting, I do not challenge Hieronymi's claim that blame is the sort of thing which stings. But some sorts of blame do not sting: the earthquake feels nothing when we blame it for the damage to buildings.

[20] Watson raises the same concern: "But how far will this [that we dislike when others resent us] take us? It is disagreeable only when the disagreement is felt. And some may be indifferent to others' disapproval altogether" (1996, p. 238).

[21] A dedicated skeptic could re-raise the Hieronymi and Watson critiques here. Even in the case of intentionally inflicted punishment, "It is disagreeable only when the disagreement is felt. And some may be indifferent to [the punishment] altogether" (quoting Watson, 1996, p. 238, replacing "disapproval" with "punishment").

blaming in the importance they carry for our interpersonal relations, for our mutual standing. But Hieronymi's argument is too strong. She argues that "the content of a judgment of ill will can carry a certain amount of force—despite being descriptive. If it [the judgment of ill will] is true, then you no longer stand in such a relationship [i.e., one of mutual regard]" (2004, p. 124). But I have blamed those close to me, and in the vast majority of cases, those instances of blaming did not disrupt the relationships. I can even recall instances of blame which went unaddressed--and yet the friendships continued. (And I'm sure I have been the target of such blamings, hopefully infrequently.) Minor blamings need not vitiate, or even significantly affect, standings of mutual regard. Friendships, after all, can tolerate occasional wrongdoings.[22] Nonetheless, these instances of blaming do seem to have carried a sting. I would regret being blamed in such a case, even if it did not threaten my relationship. The negative reactive attitudes need not erode or even threaten to erode a standing of mutual regard.

That said, Hieronymi is correct to focus on our standings of mutual regard. The reactive attitudes arise because we expect to be treated as moral peers in light of those standings. But we do not need to appeal to the apparent fragility of those standings to locate the sting of blame. We do not need to appeal to the potential for a significant if not wholly vitiating impact on our standings of mutual regard to locate the sting of blame. Rather, the sting of blame arises because, as Strawson noted, we

[22] Repeated and persistent negative reactive attitudes might lead to the erosion of a relationship of mutual regard. But this is an exceptional case, not the normal and characteristic operation of the reactive attitudes.

are sociable creatures and so, in general, we care about how others regard us, and not just how others treat us. We ordinarily care how others see us, and this is independent of the further impact that they might have on our lives. Thus we care even about the opinions of strangers and of those living in the future, though this need not entail that we care about those opinions just as much as we care about the opinions of those more proximate.

Accordingly, the sting of the reactive-attitude sort of blame falls out of its seeing-as element.[23] It is true that all sorts of blame involve some judgment. But the seeing-as element heightens the epistemic import. In the case of resentment in particular, it isn't just that we are judged to have done wrong, though that is often a significant sting. It is that we are seen in light of--in the focusing light of--our wrongdoing. We care about how we are seen, and we want others to focus on our strengths. Thus the negative reactive attitudes have a distinctive sting.

In this sense, the blaming reactive attitudes mark a central way (and perhaps the central way) that we hold each other accountable. Earlier I spoke of Watson's attributability sense of responsibility. An agent is responsible for an action in the attributability sense when the action can properly be taken to reveal something about the agent's character. This is an aretaic sort of responsibility. Watson distinguishes the attributability sense of responsibility from an accountability sense of responsibility: an agent is responsible in the accountability sense when it is appropriate to respond to the agent with adverse treatment or negative attitudes,

---

[23] This is not to say that the behavior characteristic of resentment might not sting. There can be more than one sting!

and not merely with a judgment. For Watson, the two sorts of blame are related: "accountability blame is a response to the faults identified in the aretaic sense" (1996, p. 238). But accountability requires more than attributability, for accountability involves a distinctive sting.[24] Returning to Hieronymi, we can distinguish the sting of judgments (which she accepts) from the sting of the affect-laden attitudes (which she rejects). The sting of a judgment cannot be unfair if the judgment is accurate, she insists. Even if that is so, there might be distinctive questions having to do with the fairness of accountability in light of its distinctive sting.

Taking stock, I've laid out the core of the concept of the reactive attitudes I'm concerned with, one taken largely from Strawson's work. These reactive attitudes are our affective reactions to the quality of will manifested in agent's behaviors--in our own behaviors and others' behaviors. As is commonly accepted, I see the reactive attitudes as having as central components a propositional element tracking the manifested quality of will, a characteristic felt phenomenology, and a characteristic set of behavioral dispositions. In addition to these three components, I follow important recent work from Allais, Hurley, Macnamara, Menges, and Zimmerman in stressing a further component, one informed by work on the emotions: the reactive attitudes influence how we see the agents who act, directing our attentions and interpretations. This four-element model of the reactive attitudes

[24] The relationship between the attributability and accountability is not straightforward. As Nelkin (2015) explains, one significant dispute is whether attributability might not be sufficient for accountability, even if the two sorts of responsibility are conceptually distinct.

allows us a moderately wide-scoped account of the reactive attitudes, incorporating the negative reactive attitudes like resentment but also positive reactive attitudes like gratitude, and it provides us an attractive explanation of why we should see the negative reactive attitudes as an important and distinctive sort of blame. Now I turn to consider when an agent is an appropriate target of the reactive attitudes, i.e., when an agent is responsible.

1.3 Responsibility and the reactive attitudes

1.3.1 The Strawsonian biconditional

For Strawson, the reactive attitudes are at the core of the concept of responsibility. Although there is a cottage industry of Strawson interpretations, many Strawsonians accept the Strawsonian biconditional, which I put as follows:

> an agent is responsible (for a bit of behavior) just in case the agent is an appropriate target of the reactive attitudes (on the basis of that bit of behavior).

This biconditional claims that there is a central, normative connection between (a certain sort of) responsibility and practices like blame and praise. That connection is widely accepted. We see something like it from incompatibilists like Derk Pereboom (2014), revisionists like Vargas (2013), and compatibilists like John Martin Fischer and Mark Ravizza (1998), and we see something like this from response-dependence theorists like Wallace (1994) and from realists like David Brink and Dana Kay Nelkin (2013).[25] And this biconditional can also be seen as marking the

[25] Fischer sometimes suggests that responsibility for wrongdoing and blameworthiness are not tied in this fashion, as in his (2014) review of Pereboom's book. But in those cases, it is not clear that it is responsibility that is being

relationship between blame and responsibility picked out by Watson's accountability sense of responsibility, the sort of responsibility required for us to be disposed to react to others on the basis of how they meet, exceed, or fall short of our moral expectations.[26]

But "appropriate" is a normative notion, so we need to get clear on the right normative sense. As Wallace remarks, "appropriate" is a "bland and noncommittal term[] of generalized appraisal" (1994, p. 92). I identify the sense of appropriateness at issue as fittingness: an agent is responsible (for bit of behavior) just in case the agent is a fitting target of the reactive attitudes (on the basis of that bit of behavior). And I understand the sense of fittingness here to be the sense discussed by Justin D'Arms and Daniel Jacobson in their work on the appropriateness of the emotions, especially their essay "The Moralistic Fallacy" (2000).[27] As D'Arms and Jacobson explain, emotions mark the elements of the world as having certain features. Fear, for example, marks its object as being threatening.

---

considered rather than reasons-responsiveness. That is, it seems there that Fischer is saying that reasons-responsiveness and wrongdoing are not sufficient for blameworthiness. Fischer elsewhere advocates a view of responsibility where reasons-responsiveness is one subpart of a complex account.

[26] Sometimes Watson writes as if the readiness to treat harshly is central to the sort of blame at issue in accountability. For example, he writes that the "blaming attitudes involve a readiness to adverse treatment" (1996, p. 239). If this is to suggest that one is not blaming if one is not ready to treat adversely, then I reject the suggestion. I see the readiness to treat adversely as a common element of the blaming reactive attitudes, not a necessary feature of them.

[27] While D'Arms and Jacobson's work is duly popular and is often cited to explain the sense of fittingness involved in the Strawsonian biconditional (see, e.g., D. Justin Coates and Neal Tognazzini (2016)), the idea that we can make sense of the appropriateness of our responses and reactions in fittingness terms is not new to them. We see it, for instance, in the work of Adam Smith (2010).

We can then make sense of the fittingness of the emotions by asking whether their targets in fact have the corresponding features. At least in ordinary cases, it is fitting to fear a threatening storm, but it is not fitting to fear a balmy afternoon. The storm might cause you harm, but the balmy afternoon almost certainly will not. In the case of the reactive attitudes, we can point to the content of the attitudes. Resentment is a fitting reactive attitude, i.e., an agent is blameworthy, when the agent's behavior manifests ill will, and gratitude is a fitting reactive attitude, i.e., an agent is praiseworthy, when the agent's behavior manifests good will.[28]

Fittingness is not all there is to appropriateness, especially to all-things-considered appropriateness. A reaction can be fitting though there is reason to resist, suppress, or avoid the reaction; and there can be reason to induce a reaction even though it is not fitting. It might be useful sometimes to experience a reaction in order to fit in or to share a social experience. It might be useful, for instance, to induce a feeling of fright during a horror film, even if there is nothing threatening about the film. Nonetheless, fittingness does have practical import. Centrally, the fittingness of a reactive attitude provides a pro tanto basis for the reactive attitude being practically appropriate. Something like this is widely assumed in the responsibility literature. McKenna, for instance, thinks that we must understand the sense of appropriateness in the Strawsonian biconditional as "offer[ing] a *pro tanto*

[28] That the fittingness sense of appropriateness makes reference to quality of will does not resolve the debate between quality of will and fair opportunity theories of responsibility, for much hangs on the conditions of manifestation.

reason" (2012, p. 36, emphasis in original).[29] We can see this relationship between fittingness and practical appropriateness in our explanations of why we engage in any particular instance of blame, by appealing to fittingness.

That there is this connection between the fittingness sense of "appropriate" and the practical sense of "appropriate" is a substantive and potentially controversial position, and defending it would be an important part of a full defense of a Strawsonian compatibilist account. Although that question is far broader than my current project, I can gesture at some of the elements that might ground the connection between fittingness and practical reasons by pointing to the ways that the reactive attitudes might be thought to mark an important part of healthy human life.[30] If the reactive attitudes are an important part of healthy human life, then that we have good reason to engage in the practice might be seen as providing an explanation for why we have reason in particular cases to engage in the reactions.[31]

As Fischer and Ravizza write, life without the reactive attitudes would be "cold and alienating--and highly unattractive" (1998, p. 4). A susceptibility to the reactive attitudes might be thought necessary or at least important for certain sorts

---

[29] McKenna's point is primarily that the reason be seen as a pro tanto reason rather than an all-things-considered reason. But that the relevant distinction is between pro tanto and all-things-considered reasons, not between offering a practical reason and serving some other role, shows how palatable it is to assume that responsibility for wrongdoing supplies a pro tanto reason to blame.

[30] For similar arguments, see Macalaster Bell (2013).

[31] That is, an external justification for the practice might explain why fittingness gives rise to a pro tanto practical reason, an internal justification. Consider the sort of argument John Rawls gives for the different levels of justification in his "Two Concepts of Rules" (1955). I thank David Brink for pushing me to make clearer the relationship between the different levels of justification.

of valuable human relationships.[32] It is via the reactive attitudes that we experience ourselves and others as agents and not as 'mere' causes or sites of behavior. This experience of someone as an agent is the stuff of significant interpersonal relationships. Because the reactive attitudes are the way we experience ourselves and others as agents, the reactive attitudes are an important part of how we engage in significant interpersonal relationships.

And the reactive attitudes provide important behavioral incentives.[33] We like being the objects of praise, and we dislike being the objects of blame. Because we are motivated to avoid blame, the possibility of blame can help us avoid running afoul of our own and others' moral demands. These are, of course, only imperfect mechanisms. Because interpersonal blame depends upon the interpersonal perception of wrongdoing, and because perception is a matter of appearances, the behavioral incentives are most directly sensitive to the appearance of wrongdoing. This can yield a motivation to cover up wrongdoing, and this can yield a motivation to act immorally where others have errant moral beliefs. Moreover, our concern with others' resentments is highly variable. But because we are generally motivated

[32] Most Strawsonians offer some version of this claim. We see the stronger, necessity claim in Strawson, and we see a somewhat weaker version in Wallace. But not everyone accepts the connection between the susceptibility to the reactive attitudes and centrally important human relationships. Importantly, skeptics about the applicability of the reactive attitudes such as Pereboom and Per Milam (2014, 2016) claim that we might maintain our ordinary relationships or at least comparably valuable replacements even if we abandoned the reactive attitudes.

[33] Not surprisingly, finding forward-looking import in the reactive attitudes is more common in revisionists such as Vargas (2013).

by our concerns for how others see us, we should see the reactive attitudes as having a moral-coordination effect.

The reactive attitudes also play important expressive roles. They allow us to express praise and censure.[34] Each of the various reactive attitudes are closely associated with characteristic forms of expression--verbal, facial, and otherwise. We can wince or frown, for instance, upon the appreciation of some bit of wrongdoing. Thus the reactive attitudes provide a mechanism for us to express our judgments about the ways that agents' actions reflect upon them. But this is not to say that the reactive attitudes are merely a neutral medium for the conveyance of the proposition that the target agent has missed, met, or exceeded some moral expectation. The affective component of the reactive attitudes adds a particular kind of force to the message. As Miranda Fricker (2016) insists, the reactive attitudes are emotionally laden expressions, and that emotion is significant for us. That emotion marks the reactive attitudes as a distinctive medium for communication. Moreover, plausibly, the seeing-as element helps enable our expressive behavior. By simplifying our assessments of each other, the reactive attitudes enable us to offer expressions which are more communicatively forceful because less qualified and therefore less milquetoast. Even if the milquetoast or the qualified assessment would be our all-things-considered assessment, that expression might lack

[34] Almost everyone writing about the reactive attitudes explicitly acknowledges that the reactive attitudes can express censure, and many also accept that the reactive attitudes can express praise. For many, this expressive role is the most important role that the reactive attitudes play.

communicative power, and so the focusing provided by the reactive attitudes helps us to express something more meaningful.

Finally, the reactive attitudes are valuable for us because we have limited attention but live in a complex world. As David Zimmerman explains, the emotions help us manage our attentions. He writes that "logical considerations alone, even when supplemented by any of the well-recognized epistemic and methodological principles, do not determine salience, what to attend to, what to inquire about in contexts in which this kind of focal efficiency is crucial to the advancement and protection of basic interests" (2001, pp. 534–535, punctuation and citations omitted). These points extend to the reactive attitudes. We need to attend to our own and to others' quality of will, but we cannot organize those attentions on purely principled grounds.[35] Instead, the reactive attitudes provide us a way to react to others' manifested quality of will in the face of our necessarily incomplete circumstances. They are highly important heuristics for our attention.

All of these are contingent, empirical claims, and all of them might be contested. Moreover, even if the reactive attitudes are generally beneficial, that does not mean that any of us have reason to experience them in any particular instance. We should want a fuller explanation of the connection between the practice and the instance. But these contingent, empirical claims make plausible that we have good reason to be the sorts of creatures who are susceptible to the reactive attitudes.[36]

---

[35] Although I take this denial to be true, I do not defend it here.

[36] One potentially promising defense of the reactive attitudes might point to a virtue-theoretic defense of the role of the reactive attitudes in the good human life,

And that claim makes plausible that there is a connection between matters of fittingness and practical matters.

Even supposing that the fittingness of the reactive attitudes grounds the practical appropriateness of the reactive attitudes, the fittingness of the reactive attitudes is not sufficient for the all-things-considered practical appropriateness of my particular experience of some reactive attitude.[37] Fischer and Ravizza, for instance, distinguish between blameworthiness and it being justified or appropriate, all things considered, to actually have any reactive attitude toward an agent (1998, p. 7). As Fischer later explains, "an agent can be morally responsible, but circumstances may be such as to render praise or blame unjustifiable" (2004, p. 158). And McKenna explains that, while blameworthiness can make for a pro tanto reason to blame, it does not by itself provide an all-things-considered reason to blame (2012, p. 36).

We can see this distinction between fittingness and all-things-considered appropriateness by looking at cases where facts about the agent who might

---

one consistent perhaps with a contemporary virtue theory of the sort that Philippa Foot (2001) has offered, and one evocative of Vargas's (2013, p. 172) two-level justification. Those sympathetic to virtue theories of this sort will hopefully find this suggestion illuminating. In any case, it is a placeholder for a future exploration.

[37] Is fittingness necessary? It's unclear whether responsibility for wrongdoing provides the only pro tanto reason to blame. Imagine, for instance, the sorts of reasons a utilitarian might offer to blame. If those reasons are good reasons to blame, then blameworthiness marks merely one of several pro tanto reasons to blame (even if it seems to be a privileged reason). It's also unclear whether such non-blameworthiness reasons give us reasons to blame or merely reasons to faux blame. I read the disagreement between the incompatibilism urged by Pereboom (2014) and the revisionism urged by Vargas (2013) as in large part a disagreement about this matter.

experience the reactive attitudes seem to make the reactive attitudes inappropriate all-things-considered. For instance, there is something intuitively amiss in cases where one wrongdoer blames another wrongdoer for a wrong in which they are both implicated, especially if there is not at the same time any self-directed blame.[38] As McKenna explains, "If Joe is an adulterer, then even if Josephine is blameworthy for her act of adultery, Joe might not have moral standing to hold Josephine morally responsible for her transgressions" (2012, p. 28). And we might think that this lack of standing is what makes it inappropriate for Joe to blame Josephine despite her being blameworthy. There are a number of plausible explanations for why unclean hands might defeat the appropriateness of blame. Perhaps it is that you cannot appropriately blame someone for failing to measure up to some expectation when your own behavior reveals that you yourself reject that expectation, or perhaps it is that you cannot appropriately blame someone when you have played a role in their wrongdoing.[39] Whatever the best explanation, the unclean-hands cases are one sort of case where it is not appropriate for some agent to blame a wrongdoer despite the wrongdoer being responsible for their wrongdoing and therefore blameworthy.

There are other cases where it might seem intuitive that blaming is all-things-considered inappropriate despite the agent being blameworthy. For instance, in particular cases, blaming could cause harm sufficient to outweigh

---

[38] Wallace (2011) offers a Strawsonian explanation of why it might be inappropriate to blame a blameworthy agent if you're also blameworthy: doing so might seem to suggest that you don't treat all people equally. See also Matt King (2015) for a nice, recent treatment of unclean-hands and the issue of the standing to blame.

[39] For an alternative explanation of the hypocrisy cases which appeals to the seeing-as element of blame, see my "Paying Attention to Standing" (n.d.).

blameworthiness.[40] As McKenna explains, the pro tanto reason to blame provided by blameworthiness "could be defeated by other weighty reasons" (2012, p. 36).[41] McKenna gives an example where blaming someone would cause the destruction of the planet. That consequence is so grave that we would have good reason not to blame even if the agent were blameworthy. There are perhaps more ordinary cases of outweighing. Consider a case where blaming a responsible wrongdoer would cause tremendous harm to the wrongdoer's innocent family members. This could be emotional harm, such as the harm caused to the parent as the child is blamed, or it could be more ordinary harm, such as the harm a child might suffer if blaming the parent disrupts the parent's caretaking of the child. These harms are contingent side effects of blaming, neither conceptually necessary to blaming nor (presumably) the ordinary role or purpose of blaming. But blaming could result in such harms, and there could be cases where the scale of such harms renders it inappropriate to blame the agent. In such cases, the reasons to blame given by blameworthiness are outweighed by the reasons to refrain from blaming arising from the costs of

[40] We should be careful to distinguish the consequences of blaming from the consequences of the expression of blame. That said, the expression of blame is often one ordinary consequence of blaming.

[41] The term "weighty" might suggest a particular kind of aggregation of reasons, perhaps even consequentialism about the reasons involved. I mean here to be more general than that. For now, I only need that the reasons involved can combine in some way where it is possible that there could be a reason to blame of some sort but that, at the same time, there is not all-things-considered reason to fully blame. This could involve consequentialist aggregation, but it could also involve exclusion, estoppel, or any number of other sorts of relations between practical reasons.

blaming.[42] Whether the harms of blaming outweigh the reasons of blameworthiness in any particular case will depend both upon the harms of blaming in that case and upon the force of the reasons provided by blameworthiness in that case. But in the cases where the harms do outweigh blameworthiness, even though the agent is blameworthy, it is not appropriate to blame the agent.

It is important to keep these wide-ranging defeaters of fittingness in mind in order to avoid the moralistic fallacy identified by D'Arms and Jacobson. D'Arms and Jacobson offer the example of a nasty joke. We might judge that it would be wrong, all-things-considered, to laugh at the joke, and this judgment could lead us to infer that the joke is not funny. However, as D'Arms and Jacobson explain:

> to commit the moralistic fallacy is to infer, from the claim that it would be wrong or vicious to feel an emotion, that it is therefore unfitting. We shall contend, to the contrary, that an emotion can be fitting despite being wrong (or inexpedient) to feel. In fact, the wrongness of feeling an emotion never, in itself, constitutes a reason that the emotion fails to be fitting. (2000, p. 69)

We must be careful distinguish the reasons which tell against an emotion from the reasons which make the emotion unfitting. And this lesson about the emotions can shed light on how we should think about the reactive attitudes. The phenomena which can bear on the all-things-considered appropriateness of the reactive attitudes are wider than the phenomena which bear on the fittingness of the reactive attitudes. Failing to keep this in mind can lead us to misunderstand the

---

[42] These harms need not be harms imposed upon a third party. As the forgiveness literature makes clear, blaming can be bad for the agent who blames. It can be upsetting and disruptive, and it can distract the blaming agent from focusing on more productive matters. These costs might give an agent good reason to forego a fitting blame response in order to protect his own interests.

nature of the reactive attitudes, for we might infer from the fact that it is wrong or vicious to feel a reactive attitude that the reactive attitude is unfitting, and we might then infer either that the reactive attitude does not correctly present the world's evaluative features to us or that fittingness has to do with things beyond correct presentation; neither inference would be warranted on such grounds. Accordingly, while my primary subject is responsibility and the fittingness of the reactive attitudes, it will be important to return regularly to the defeaters of fittingness. Being aware of the presence of defeaters will help preclude committing the moralistic fallacy, and the defeaters are important in their own right for understanding the norms of blaming.

1.3.2 Realism or response-dependence?

Much of Strawson's account of moral responsibility is widely accepted--in particular the revival of interest in the reactive attitudes and the connection between blameworthiness and responsibility. However, some balk at Strawson's apparent further commitment to the idea that moral responsibility is, in its essence, about our responding practices, such that the contours of moral responsibility depend upon the contours of our responses or our responding practices. In contrast to such apparently response-dependent views, some of these critics favor realism, claiming an independent status for moral responsibility. This distinction between realism and response-dependence about responsibility is often taken to be a foundational question. My focus on the reactive attitudes might suggest that I am committed to a response-dependent view. However, while we can learn much from

the response-dependence/realism debate, rejecting what I will call performed response-dependence and accepting what I will call methodological response-dependence, I prescind from taking any stand on the ultimate question, suggesting that the stakes are smaller than many have thought.

When thinking of moral responsibility, there are different views which we might call response-dependent views.[43] One of those views ties moral responsibility to the ways that we in fact do respond to each other, the performed patterns of behavior. I call these views performed response-dependent views. One example of a performed response-dependent view might be this view:

> an agent is responsible (for bit of behavior) just in case holding that agent morally responsible (e.g., praising or blaming) fits the general patterns of praise, blame, and other reactive responses which mark the relevant society.

Imagine a society where it is normal to blame even the youngest children when they do wrong but where those who suffer from certain mental illnesses are not held responsible. In that society, a young child would in fact be responsible, because that fits the general pattern of praise and blame. However, not everyone is responsible in that society: those suffering from certain mental illnesses would not be responsible.

This view provides a way to make sense of certain sociological behaviors. However, it does not provide a way for evaluating whether a widespread practice is in error (in fact, it denies that possibility), nor does it provide a way to determine

---

[43] Here I look at views which might be thought of as "pure response-dependent views." But those are not the only views which we might imagine. On one reading, McKenna (2012) offers a blended view, with both realist and response-dependent elements.

whether a society's reactionary practices have become better or worse. So while I allow that there might be reason to be interested in a more sophisticated version of the performed response-dependent view, it is not the view I'm interested in, and so I set it aside.[44]

Then there is methodological response-dependence. On this view, we can appeal to our judgments about instances of appropriate response as evidence in our inquiry about responsibility. Our judgments about responsibility are dependent upon the evidence we gather from our actual responses, our intuitions regarding cases, and the like. So long as we accept the Strawsonian biconditional, and so long as we take our judgments and intuitions about particular cases as generally reliable, we can use our intuitions about responses to pick out the contours of responsibility. Methodological response-dependence does not entail that our responses or the appropriateness of the responses are explanatorily or metaphysically prior to the facts about responsibility, though of course it is consistent with these priorities. Rather, methodological response-dependence only requires that we might at least sometimes have readier access to our intuitions about the appropriateness of certain responses than to the facts about moral responsibility. While this is a comparably innocuous claim, it reflects a shallow response-dependence, perhaps

[44] For similar reasons, I set aside any view which grounds responsibility in the reactions which we are inclined or disposed to have.

not even deserving of the name. Realists can and do accept what I'm calling methodological response-dependence without abandoning their realism.[45]

Accordingly, I set aside performed response-dependence, and I accept methodological response-dependence. But this still leaves open what seems to be the core of the dispute between realists and response-dependence theorists. Begin by attending to the Strawsonian biconditional. Recall that the Strawsonian biconditional claims that an agent is responsible for some bit of behavior if and only if the agent is a fitting (i.e., appropriate, in the relevant sense) target for the reactive attitudes with respect to that behavior. But we might then ask which side of the biconditional has explanatory priority. Is it that the agent is responsible because the agent is a fitting target of the reactive attitudes, which is the response-dependence theorist's claim? Or is it that the agent is a fitting target of the reactive attitudes because the agent is responsible, which is the realist's claim?

Brink and Nelkin (2013) argue that even this sort of response-dependence is unable to ground criticism of our extant practices. It's not clear why this should be so. As Wallace makes clear, this sort of response-dependent view can make sense of the truth conditions of the appropriateness of the reactive attitudes. On Wallace's response-dependent view, for instance, it is appropriate to resent a wrongdoer (thus making the wrongdoer responsible) if, only if, and because the agent

---

[45] If this is the sort of position response-dependence theorists are claiming, then they are in essence denying any significant difference between realism and response-dependence. Moreover, if this is the only sense in which a theory is response-dependent, it is hard to see why we should not call such a theory a realist theory.

possessed the two reasons-responsiveness capacities at the time of the wrongdoing. If we hold responsible those who lack the two capacities, we are in error, even if it is our society's practice to do so or we are inclined to do so. That Wallace claims that his view can support criticism of our extant practices might suggest that, contrary to his self-description and usual reception as a response-dependence theorist, Wallace is best understood as a realist. But Wallace seems committed to his response-dependence: "I cannot see how to make sense of the idea of a prior and independent realm of moral responsibility facts. … it seems incredible to suppose that there is a prior and independent realm of facts about responsibility to which such emotions and actions should have to answer" (1994, p. 88).

Here is one way to make sense of how Wallace could maintain his response-dependence line and yet offer a theory which can support criticism of our practices. Notice that Brink, Nelkin, and Wallace might all accept both of the following propositions:

> (1) an agent is responsible if and only if the agent is reasons-responsive; and
> (2) an agent is responsible if and only if the agent is a fitting target of the reactive attitudes.

For the response-dependence theorist, (1) is a synthetic claim, and (2) is an analytic claim, and (2) must be appealed to in explaining (1). As Neal Tognazzini understands the position of the Strawsonian response-dependence theorist:

> the question will always remain: what is it about those capacities that 'calls for' or justifies blame? Add as much to the list as you like, but as Strawson says, 'there still seems to remain a gap between its applicability in particular cases and its supposed moral consequences.' (2013, p. 1302, quoting Strawson)

By contrast, realists would claim that (2) is a synthetic claim, even if different sorts of realists might accept different characterizations of (1). To characterize those different ways of thinking about (2): for the realist, that responsibility grounds the fittingness of the reactive attitudes is a property (perhaps a morally necessary property) of responsibility, whereas for the response-dependence theorist, that responsibility grounds the fittingness of the reactive attitudes is the essence of responsibility.[46]

But that the debate between realists and response-dependence theorists might turn on the distinction between essences and properties shows just how incredibly close realists like Brink and Nelkin are to response-dependence theorists like Wallace.[47] Indeed, on this construal, some self-described realists might be best seen as response-dependence theorists. If the best understanding of their view makes fairness essential to the concept of moral responsibility, then the thing that must be fair is likewise intimately part of the concept--but that thing is the response! This might be the sort of thought which leads, Fischer, ostensibly a realist, to write that "there does not seem to be any interesting difference, with respect to the issue of irreducible normativity!" (2012, p. 141). We should not so quickly concede the matter. For example, realists need to identify the sort of realism they

[46] Although the distinction between essences and properties is ancient, found for instance in Aristotle, for a recent treatment of the distinction between essences and properties, see Kit Fine (1994).

[47] See Patrick Todd's (2016) argument about how difficult it is to determine exactly what the various participants in the debate between the realists and the response-dependence theorists are taking themselves to claim.

are staking out. A realist who identifies moral responsibility with the manifestation of quality of will and a realist who identifies moral responsibility with the reasons-responsiveness capacities disagree about the concept at issue. Their disagreement is largely a disagreement about vocabulary; they are simply talking about two different notions of responsibility. The substantive questions are downstream--which sorts of responsibility play which roles. By contrast, a response-dependence theorist will see the choice between manifestation theories and capacity theories as a substantive choice about responsibility. Still, while there are surely stakes to this debate, I sidestep the question. As Vargas concludes, "Rather than trying to answer the 'is it realism?' question, we are better off focusing on more fine-grained questions about successful reference, truth-functionality of a given bit of discourse, licensed inferences, and the status of imputed properties" (2013, p. 123).

Chapter 2 The reasons-responsiveness account of moral responsibility

In this chapter, I offer a reasons-responsiveness account of moral responsibility. I claim that an agent is morally responsible in the sense that renders the agent a fitting target of reactive attitudes like praise and blame if and only if the agent was reasons-responsive at the time of the behavior at issue. And an agent is reasons-responsive if and only if the agent possesses the capacity to respond to reasons. This is a capacity compromised of many component agential capacities. As H. L. A. Hart explained, "[t]he capacities in question are those of understanding, reasoning, and control of conduct: the ability to understand what conduct legal rules or morality require, to deliberate and reach decisions concerning these requirements, and to conform to decisions when made" (1968, p. 227). This reasons-responsiveness account provides a satisfying explanation of the patterns of our reactive attitudes, and it provides a satisfying sense of control. It also helps us understand the relevance of other matters for responsibility, matters like the role for empathy or the presence of especially strong urges. To see all this, and to have the materials needed to assess the role of history, I offer a fuller development of the reasons-responsiveness account of moral responsibility here.

2.1 Explaining our paradigmatic patterns

I expect that we are pretty good at ascribing responsibility. Because of that, we can pick out a plausible candidate account of moral responsibility by looking at the patterns of our ascriptions of responsibility. By referencing considered

judgments about cases both real and hypothetical, we can contrast paradigmatic cases of responsibility with paradigmatic cases of irresponsibility and with paradigmatic cases of compromised responsibility. Those comparative patterns provide good if defeasible evidence of an underlying account of the conditions of moral responsibility.[48]

The paradigmatic responsible agent is an informed, mature, deliberative adult who reflects upon her options, makes a decision under no particular pressures, and then acts on that decision. When such an adult acts wrongly, we usually blame her. By contrast, we are less likely to blame young children or the sufferers of certain sorts of mental illness. When a child acts wrongly, we correct, remonstrate, and teach, but we do not ordinarily blame. We might come to believe that a child's behavior manifests ill will, but we are unlikely to experience the phenomenological responses characteristic of resentment, we are unlikely to respond with the characteristic behaviors, and we are unlikely to frame our perceptions of the child in light of the wrongdoing. Likewise, when someone

[48] A full defense of the reasons-responsiveness account of moral responsibility using this method should involve significant social science. Here I rely primarily upon my own intuitive judgments about cases and my own intuitive sense of the patterns of social responses, all very much defeasible evidence. I take my own reactions to be prima facie evidence of the social phenomena (recognizing the hubris that requires). I do not take my own reactive attitudes to be particularly idiosyncratic. Alternatively, I take this work to be an investigation of what I in particular mean by moral responsibility. On that latter read, when I make claims about what we do or are inclined to do, those are best understood as claims about what I do or am inclined to do conjoined with claims about my rough sense of what we (that group itself understood but gesturally) do or are inclined to do.

suffering from certain forms of mental illness commits a wrongdoing, we quarantine, treat, and sympathize. There, too, we do not ordinarily blame.

The rivals to the reasons-responsiveness theories include characterological theories and libertarian theories. The characterological theories claim (roughly) that an agent is responsible for those behaviors which reflect their character.[49] A full assessment of a characterological theory would require full expositions of the ideas of character and of reflection. However, supposing that ordinary virtues and vices at least partly constitute character, we might then think that an agent is blameworthy if her wrongdoing reflects vice and that an agent is praiseworthy if her behavior reflects virtue.

But such characterological theories do not seem well-suited to explain our intuitive judgments. We might be confident assessing that an informed, deliberative adult who regularly does wrong and takes unconflicted pleasure in her wrongdoing is a paradigmatically vicious agent. The theory might then rightly hold this adult agent blameworthy. But we should not assume that children and the mentally ill cannot likewise be marked by deeply vicious characters. Children can be greedy, vengeful, and malicious. Even if character requires a certain sophistication, such that it makes little sense to think of very young children as having any character at all, and even if young children's characters are fluid and developing, I can imagine a child young enough to make the reactive attitudes intuitively inappropriate but old

[49] Although Vargas is not a characterological theorist, his recent discussion (2013, pp. 133–157) lays out the theory (and several variants) clearly and offers fair treatments of the theory's strengths and weaknesses.

enough to have and act upon distinctive character traits. The characterological theory would seem to inappropriately hold this child responsible. Likewise, there's little reason to think that the mentally ill universally lack character. Mental illness might sometimes occlude character, and it might sometimes mold and affect character; but mental illness does not always obliterate character. In fact, in some cases, we might think that certain sorts of mental illness affect or even create character traits. If so, we might expect that some mentally ill agents intuitively entitled to excuse nonetheless act upon character traits. While these claims are tentative, it is intuitively plausible that all these sorts of agents--adults, at least some children, and at least some of the mentally ill--can possess and act upon character, and so character does not seem to be what distinguishes the intuitively responsible from the intuitively not responsible.

There is a second class of cases which seem to trouble the characterological theories: cases of agents who are intuitively responsible despite acting out of character. That an agent is honest does tell us something about the agent's propensity to lie, but even an honest agent might nonetheless sometimes lie. Indeed, I think (perhaps optimistically) that much of the wrongdoing for which we hold each other responsible is just this sort of out-of-character behavior. But the characterological theory would seem to require us to either revise our judgments regarding character attributions significantly or to let all of these intuitively responsible agents off the hook.

The libertarian fares no better in explaining the patterns in our judgments of paradigm cases. A central libertarian thought is (roughly) that an agent is responsible for a bit of behavior only if the agent acted while at liberty to do otherwise.[50] The libertarian theories do not seem to offer any promising distinction between the paradigmatic cases. Often the thought is that an agent is not at liberty if the agent's behavior is constrained by the past and the laws of nature. However, the laws of nature do not distinguish between the paradigmatically responsible adult, the young child, or the sufferer of mental illness. In all three cases, it seems that the person's choices are the products of their past, so in all three cases it seems that the libertarian would be led to deny responsibility.

The libertarian cannot remedy this by appeal to the lay sense of the possibility of doing otherwise. It is ordinary to imagine an adult wrongdoer having done otherwise. We commonly imagine someone having chosen otherwise, and we commonly imagine what might have resulted from alternative choices. Even if those thoughts turn out to be misleading or merely epistemic, this is a normal, lay sense of the possibility of doing otherwise. But this sense won't distinguish our cases, because we can imagine a child wrongdoer having done otherwise, and we can imagine an agent beset by mental illness having done otherwise (though our

[50] For thoughtful defenses of the libertarian position, see Alvin Plantinga (1974), Randolph Clarke (1993), and Robert Kane (1996). Some libertarians are global incompatibilists, thinking that determinism robs all of us of responsibility. That argument is usually grounded in the attractiveness of principles about control and the like, not in the ability to explain the appropriateness of the general patterns of our behavior.

imaginations here will be informed and limited by our understanding of the particular mental illness).

Neither characterological theories nor libertarian theories can readily explain both why adults are ordinarily blameworthy and children and the mentally ill ordinarily are not.[51] Instead, we can make that distinction by considering the various actors' normative capacities. The adult's normative psychology is such that it is reasonable to expect that, as Fischer and Ravizza explain, "There would be a tight fit between the reasons there are and the reasons the agent has, the agent's reasons and his choice, and his choice and his action" (1998, p. 42). The adult agent's reasons-responsiveness capacities ground this expectation. By contrast, the young child has still-developing capacities for responding to reasons. This explains why the child often makes practical and moral mistakes. We might also expect at least some mentally ill agents to have compromised or occluded capacities for responding to reasons, and this could explain why those agents might make practical and moral mistakes. Accordingly, it is the capacity to respond to reasons--

[51] Of course, defenders of these theories might either seek revision of our intuitions or offer more complex defenses of their theories. I do not mean to offer a final refutation of either sort of theory here. However, I suspect that if either theory remedied these problems, it would most likely result in severely reducing the gap between that theory and the reasons-responsiveness theory. For instance, the characterological theory might claim that only certain expressions of character count--perhaps just those which arise in the right agential way. But that comes awfully close to requiring the possession of the agential capacities, even if the starting point differs. In any case, I leave the final addressing of these competitors for later.

and not just the general capacity to do otherwise--which best explains our patterns of assessing moral responsibility.[52]

Because the normative capacities play a central role in the reasons-responsiveness theory of moral responsibility, the theory should offer an analysis of capacity. Here, I appeal to an inchoate but intuitive notion of an agential capacity. It seems intuitive that I now have the capacity to read. That provides part of the explanation of how I read on any particular occasion. I have this capacity even when I'm not exercising it, and I have the capacity even when it seems I could not exercise it (e.g., while I'm sleeping or while I'm sitting in the dark). The capacity to read is something that I might sometimes exercise intentionally, deciding to read a paper and then doing so, for instance, but it is also something that I might exercise without any intention to do so, reading the name of a store from a sign I happen to notice when I glance outside. This, roughly, is the sort of agential capacity at issue (though the capacity to read is a relatively high-level capacity, and it does not capture all of the features which appear in the wide range of agential capacities). We could scarcely understand agency without employing the notion of such agential capacities.

The possession and exercise of such intuitive capacities is consistent with physical determinism. My history and the laws of nature might combine to make it so that I lack the opportunity to exercise some capacity, or they might make it so

---

[52] Many philosophers have espoused reasons-responsiveness accounts of moral responsibility. Wolf (1990), Wallace (1994), Fischer and Ravizza (1998), Nelkin (2011), and Brink and Nelkin (2013) have been particularly influential.

that I decline to exercise some capacity. But it would be significantly revisionary to say that determinism makes it so that I lack capacities altogether. Plausibly, that we do something implies that we can do that thing, and that we can do something implies that we have the capacity to do it. Thus, if we rejected capacities in accepting determinism, it would not be clear how we could explain that I do what I do, or at least the explanations would look quite alien to the explanations we ordinarily offer, and so determinism would be incompatible with the ordinary explanation of our actual behavior, not merely with moral responsibility. This is too revisionary, and so we should be able to accept both physical determinism (though we need not be committed to it) and the presence of the capacities for action. This latter gives the compatibilist an important argument for her compatibilism.

I sidestep important questions about what grounds those capacities. Plausibly, the capacities are grounded in features of the agent's psychology, and, plausibly, those features plausibly bear important relationships to structures within the brain.[53] Fully developing the reasons-responsiveness account will require identifying the nature of that grounding and those contours of those relationships. That development would be an important part of identifying the diagnostic tools to look for the capacities in particular agents. But I leave those questions for later.

---

[53] This is not to say that we should expect the capacities to be identified with particular brain structures. The capacities might well involve psychological phenomena not reducible to brain structures, and they might well involve social phenomena.

2.2 The component capacities

We can be more precise about the reasons-responsiveness capacities which ground moral responsibility by looking at the component capacities which comprise reasons-responsiveness. We expect the adult to conform her behavior to relevant reasons. But in order to comport her behavior to the relevant reasons, the adult needs to be able to assess the relevant reasons, and the adult needs to be able comport her behavior to her normative assessments. Accordingly, like virtually all reasons-responsiveness theorists, I identify two general component capacities, one cognitive, one volitional.

The cognitive component capacity is the capacity to recognize the presence and practical relevance of moral reasons. One significant reason we excuse children from responsibility is their inability to understand the significance of their behavior. And when we excuse the insane, it is often because their illness compromises their ability to make good sense of their circumstances. In the case of the paradigmatically responsible adult, by contrast, we expect the agent to understand both the options they face and the significance of those options.

Some distinguish this cognitive capacity from a separate epistemic constraint on responsibility. For instance, following Aristotle, Fischer and Ravizza distinguish the excusing force of ignorance from the excusing force of the lack of control, where the cognitive capacity plays a role in the agent's having control over her behavior (1998, p. 13). For Aristotle, Fischer, Ravizza, and many others, innocently not knowing that you are about to miss a good friend's birthday party and knowing that

you are about to miss the party but not being able to see that as a significant omission are thought to be different grounds for excuse. In the one case, you are innocently ignorant of some descriptive fact, and in the other case you are innocently ignorant of some normative fact. I refrain here from deciding whether these are two different sorts of excuses or two variants of the same cognitive excuse.[54] On my understanding, what matters is whether the agent has the capacity to discern the normatively relevant options. That involves both discerning the options and seeing their significance.

The cognitive capacity requires that the agent be capable of discerning moral reasons in particular. An agent who could respond to pragmatic but not moral reasons would not be a responsible agent.[55] On a lay understanding, some psychopaths are like this.[56] Plausibly, very young children are sometimes like this as well, being far more sensitive to their own interests than to interests of others. This contrast in sensitivity might suggest that very young children are sensitive to

---

[54] The account could also be complicated to distinguish between excuses grounded in cognitive incapacity and excuses grounded in ignorance, given that an agent might both possess the cognitive capacity and nonetheless be ignorant. I am inclined toward a unitary account focused on the capacities, though this raises issues about culpable ignorance; in any case, this is an issue worth further reflection, and I thank David Brink for pushing me to think more about this.

[55] Perhaps that agent would perhaps be pragmatically responsible but not morally responsible. That is, there could be pragmatic reactive attitudes, in addition to the more familiar moral reactive attitudes, and pragmatic competence might the condition of the pragmatic responsibility required for the appropriate application of those attitudes.

[56] For thoughtful discussions of the moral responsibility of psychopaths, see Cordelia Fine and Jeanette Kennett (2004), Neil Levy (2007a, 2007b), Brink (2013), and David Shoemaker (2015).

pragmatic reasons but that they have not yet developed due sensitivity to moral reasons.

And the cognitive capacity requires more than simply "the ability to parrot the moral principle in situations in which it has some relevance," as Wallace explains (1994, p. 157). This can explain why children have reduced or no responsibility--they often cannot grasp the reasons for the import of the moral reasons they can recite. "Just because" is good enough to insist upon compliance, but it is not good enough for moral responsibility. The agent must have access to a genuine understanding of the relevant moral reasons. A genuine understanding entails the ability to apply the principle to wide range of relevant situations.[57] We might require more. We might require that the agent have access to a robust understanding of the reason at issue, one including a sense of the grounding of the principle. But we must be careful not to be too demanding. Few, if any of us, routinely act upon moral principles which we grasp down to basic fundamentals, and likely few of us have the capacities to routinely act with such deep understandings. Such understandings are far too rare to explain our judgments in the paradigmatic cases. In identifying the Goldilocks middle ground, it is important to recognize that, just like the reasons-responsiveness capacity itself, the cognitive capacity involves further component capacities.[58] The capacity to understand the

---

[57] In fact, what matters is access to a genuine understanding of how the principle applies to the particular situation. However, at least in ordinary cases, a genuine understanding will yield wide usefulness.

[58] Wallace (157-158) is particularly clear that the cognitive capacity is comprised of a number of further capacities.

relevant moral reasons requires being able to attend to the relevant factors, being able to distinguish between related features of the world, and being able to make needed moral judgments. And these subcomponents may themselves further ramify.

The second capacity, the volitional capacity, is the capacity to comport one's behavior to one's normative assessments. Another significant reason we excuse children is their impulsivity. Sometimes children act poorly even when they know or should know better. They act on an immediate want or urge, without letting that urge be mediated by what they know, in some sense, to be a better, contrary choice. Likewise, addiction can interfere with an agent's volitional capacity. The addict might be excused if his normative control is such that he would find himself indulging the addiction even in the face of an earnest judgment otherwise. And even depression might be understood as at least sometimes compromising our volitional capacity so much as to excuse, at least where it robs us of motivation.

Philosophers have made less progress exploring the volitional component than they have with regard to the cognitive component. That can explain why Fischer and Ravizza think that reactivity, their term for the volitional capacity, is "all of a piece" (1998, p. 73). Fischer and Ravizza claim that the volitional capacity is not indexed to particular reasons to act. If an agent's mechanism can react to one reason, then that mechanism can react to any reason.[59] The lack of exploration of

[59] Fischer and Ravizza predicate reasons-responsiveness of mechanisms, not agents. For them, an agent is responsible for their behavior in an ordinary case if their action is the product of a reasons-responsive mechanism. The mechanism is "the process that leads to the relevant action… the 'way the action comes about'" (1998, p. 38). An example of a mechanism is ordinary deliberation. Although I do not

the volitional capacity can also explain why there has been great reticence to include a volitional basis for excuse in the criminal law's insanity defense.[60] But I have little reason to think that the volitional capacity is any less complex than the cognitive capacity. I leave it here open whether the agent must have the capacity to act upon distinctively moral reasons and whether, in doing so, not just any causal interaction will suffice.

Seeing that reasons-responsiveness is made up of a range of further, increasingly more basic capacities has a number of benefits. It can help us distinguish different reasons warranting excuse. Even within one diagnostic category, like childhood, we can distinguish different ways that the excuse functions, sometimes cognitive, sometimes volitional. Making those distinctions can help us get our excuses right, both as to identifying which agents are entitled to excuse and as to identifying which conditions properly ground excuse. Then, seeing that reasons-responsiveness is made up of these more basic capacities can also show us that reasons-responsiveness is the right condition of moral responsibility in part

---

generally adopt the convention of talking of mechanisms rather than agents, I follow it here to avoid attributing to Fischer and Ravizza the claim that an agent's volitional capacity is "all of a piece."

[60] Stephen Morse (2002) is a particularly forceful critic of including a volitional-capacity excuse. Morse argues that excusing for loss of volitional control gains credibility because we are not careful to distinguish that loss of volitional control from the sorts of mechanical forcings which result in movements but not actions (a strong wind, for instance) or from the sorts of hard choices which might change the relative import of the choices faced. Both of these offer grounds for refraining from blame, the former because the ostensible wrongdoing was not an act, the latter because the act was not a wrongdoing. Morse then argues that there is nothing left for a volitional control element to do. Morse argues that we get all the true excuse which we might need from the cognitive element.

because it is tied to the proper functioning of agents. These capacities are among those capacities which make up agency generally. It is not just that an agent has the propensity to act differently in response to different reasons, because we might conceive of an automaton which can, in some perhaps minimal sense, respond to reasons. Rather, a well-functioning agent can respond to reasons because the well-functioning agent has these further component capacities. An actor who lacked the capacities could not be acting as an agent, and that can explain why such an actor would not be responsible as an agent.

2.3 Matters of degree

I do not here offer a full account of agential capacity; an intuitive notion is enough for my purposes. However, whatever the underlying metaphysics of the agential capacities, they should be scalar in nature. Reasons-responsiveness--like many capacities--is a matter of degree. Some agents are barely responsive, some agents are ordinarily responsive, and some agents are highly responsive.[61] This can explain some of the patterns of our responses, and being attuned to this feature of the reasons-responsiveness capacities will make it easier to account for the effects history can have on our intuitions.

---

[61] There are two different senses in which an agent might be more or less reasons-responsive. First, an agent might be reasons-responsive to more or fewer sorts of reasons. Thus an agent who is widely reasons-responsive is thereby more reasons-responsive than an agent who is insensitive to many sorts of reasons. Second, an agent might be more responsive to some particular reason than another. Thus, reasons-responsiveness is a matter of degree both with respect to scope and strength. I elide the distinction between these senses here, although there are interesting questions about both senses. I thank David Brink for pushing me to think more about this.

While the exact metaphysics of capacities might be up for debate and variable, it is intuitive that capacities can be a matter of degree.[62] We readily speak, for instance, of the capacities of ordinary objects in degree terms. "This tent can protect you from rain," we might say, and that shielding feature is a capacity of the tent. But some tents are sturdier than others, and so we understand that different tents have rain-shielding capacities of differing degrees. A flimsy tent has a minimal capacity to protect from the rain, and a technical, mountaineering tent has a great capacity to protect from the rain. We can also make sense of capacities of differing degrees with respect to our capacities for different actions. Some people lack the capacity to run a six-minute mile. Paula Radcliffe, by contrast, has the capacity to run a six-minute mile, and hers is a ready capacity. She can run a six-minute mile regularly, with little preparation, and with little commitment and effort. I'm in the middle. I am a regular runner, not a spectacular runner. I have the capacity to run a six-minute mile, but I could do so only with tremendous commitment and effort, and I could do so only irregularly. I have a moderate capacity to run a six-minute mile. These examples show several ways that we might discern different degrees in capacities: one capacity might be exercised more readily than another, one capacity might be easier to exercise than another, one capacity might succeed in broader contexts than another, etc.

---

[62] For a pithy argument suggesting that the degree of responsibility is related to but does not exactly track the degree of capacity, focusing in particular on the relevance of external circumstances, see Jesper Ryberg (2014).

We should see the component capacities comprising reasons-responsiveness as admitting of degrees.[63] Think of the cognitive capacity, the capacity to recognize the presence and relevance of moral reasons. One agent might recognize the presence and relevance of moral reasons only rarely, only with effort, and only for the most striking moral reasons. That pattern of sporadic performance gives us reason to think that the agent has but a slight cognitive capacity. Another agent, by contrast, might be readily and widely aware of the moral reasons bearing on her choices. That constant performance gives us reason to think that the agent has a genuine and strong cognitive capacity. This is likewise so for the volitional capacity, the capacity to comport one's behavior to one's assessment of the circumstances. Here we have readier vocabulary, the talk of willpower. Akratic agents have weak volitional capacities, and they are only barely (and thus usually only rarely) able to convert their assessments of their normative circumstances into corresponding behaviors. Strong-willed agents, by contrast, have great volitional capacities, and they readily (and thus often) convert their normative assessments into corresponding behaviors.

As agents can possess the component capacities to greater or lesser degrees, we should likewise see reasons-responsiveness itself as a matter of degree. Some agents, barely responsible, have but a slight capacity to respond to reasons. Some

---

[63] Although reasons-responsiveness philosophers have consistently acknowledged that the capacities at issue might well be a matter of degree, there has been little sustained attention to this issue. That said, recent papers from Coates and Philip Swenson (2013) and Nelkin (2016) offer promising investigations of the scalar nature of these capacities.

agents, highly responsible, have a great capacity to respond to reasons. Because the degree of an agent's reasons-responsiveness is a function of the degrees of his component capacities, we should expect that there could be multiple combinations which might make for roughly similar overall degrees of reasons-responsiveness. For instance, if one agent possesses only a slight cognitive capacity but matches it with a strong volitional capacity and another agent possesses a strong cognitive capacity and a slight volitional capacity, it might nonetheless be that these different agents possess reasons-responsiveness capacities of roughly the same overall degree. And the further we ramify the component capacities, the greater variety of instantiations we should expect to see of the particular degrees.

Because the reasons-responsiveness capacities come in degrees, and because responsibility is a matter of the reasons-responsiveness capacities, we might expect responsibility too to come in degrees.[64] An agent might be more or less responsible for some particular action. This means that an agent might be more or less blameworthy for the same bit of wrongdoing. That responsibility might come in degrees in just this sense fits with ordinary experience. We familiarly think that older children are more responsible than younger children.[65] As children age, their reasons-responsiveness capacities develop into mature capacities, and along this

[64] Seeing responsibility as a matter of degree is not the only option. Among other possibilities, the scalar capacities might result in binary responsibility verdicts if we have a threshold account of responsibility.

[65] See, e.g., Brink (2004), with respect to the culpability of children. In this section, I am also borrowing from Brink's current thinking about how the criminal law might account for partial responsibility.

path, children can be seen as increasingly responsible. That they are increasingly responsible is reflected in the increasing severity of the sanctions they face.

The importance of the different degrees is sensitive to the context in which we agents find ourselves. As Brink and Nelkin (2013) point out, reasons-responsiveness interacts with situational features.[66] For example, duress can compromise blameworthiness just as mental illness can. Both duress and mental illness can make it more difficult to respond to moral reasons, and so both can reduce an agent's blameworthiness when he fails to so respond. Duress is one sort of excusing situational element, but there are many others besides. Chaos, for instance, is an often-overlooked situational factor that might reasonably be thought to ground at least a partial excuse. Two otherwise similar agents, with similar capacities, who fail to respond to the same reason to do otherwise might be responsible to different degrees if the first agent failed to respond in sedate conditions, with little else demanding her attention and little interfering with her

[66] For them, the overarching responsibility concept is the fair opportunity to avoid wrongdoing. Reasons-responsiveness is one component of the fair opportunity to avoid wrongdoing, and situational factors are the other component. Brink and Nelkin's observation about the relevance of situational factors to blameworthiness might seem to threaten the equating of responsibility and reasons-responsiveness. If blameworthiness is a matter of responsibility and wrongdoing, and if the situational factors do not make the wrongdoing more or less wrong, then it might seem that the situational factors can affect blameworthiness only by affecting responsibility. If that's so, then it seems that responsibility must be more than reasons-responsiveness, at least if we accept the plausible claim that the situational factors are external to reasons-responsiveness. It's not clear that this is the only way to incorporate Brink and Nelkin's observation, and, in any case, accepting a modified understanding of responsibility along those lines would require notational changes to my argument throughout, but not substantive changes.

deliberations, while the second agent failed to respond in chaotic conditions, with his attention thinly stretched and his deliberations constantly interrupted.

Although we could see responsibility as being fundamentally a matter of degree, our everyday responsibility talk often seems binary. We can account for this feature of our ordinary practice by thinking of our responses as reflecting practical thresholds. A person is or is not responsible, in ordinary cases, and while big differences in psychology or big differences in outside conditions can change responsibility verdicts, slight differences do not.[67] A mild temptation has little effect on our responsibility assessments, for instance. Our everyday responsibility talk reflects chunky, discrete assessments of our responsibility. This should not be surprising. We are limited assessors of each others' responsibility, with limited time, limited evidence, and the like, and so practical constraints lead us to make simpler assessments than the fully scalar, continuous assessments the underlying metaphysics might admit of. As with many normative thresholds, the particular, appropriate level is a function of a number of practical concerns, but we should expect that the relevant threshold is a moderate threshold.

Fischer and Ravizza offer the most famous moderate threshold.[68] But while they are right to look for moderate reasons-responsiveness as the relevant threshold, their particular account will not do. Fischer and Ravizza distinguish three levels of reasons-responsiveness: weak, moderate, and strong. These levels are

---

[67] I thank David Brink for conversation on this.

[68] For the core of Fischer and Ravizza's argument for their moderate account of reasons-responsiveness, see (1998, pp. 62–76).

picked out by patterns of actual and counterfactual response. A weakly responsive agent[69] is one who successfully responds to relevant moral reasons in at least one actual or counterfactual situation, and a strongly responsive agent is one who successfully responds to relevant moral reasons in all situations, actual and counterfactual. A moderately responsive agent is somewhere in between. The moderately responsive agent has a weak volitional capacity: the moderately responsive agent successfully comports her behavior to her assessment of the circumstances in at least one case, actual or counterfactual. Fischer and Ravizza, however, are more demanding with respect to the cognitive capacity. There, they require that two conditions be met. First, the agent must successfully respond to relevant moral reasons in at least one actual or counterfactual situation. Second, the agent's responses must yield an understandable pattern. For example, if a reason is important enough to engender response in one case, it should likewise engender responses in most similar cases. Of course, an understandable pattern might be an imperfect pattern, so the moderate cognitive responsiveness is still weaker than strong cognitive responsiveness. For Fischer and Ravizza, then, moderate reasons-responsiveness is weak reasons-responsiveness combined with an understandability constraint on the cognitive capacity.

This asymmetry presents a puzzle. If both component capacities are a matter of degree, why be more demanding of one than of the other? The best interpretation of Fischer and Ravizza's moderateness is as evidentiary. What they ultimately care

[69] Again, Fischer and Ravizza predicate reasons-responsiveness of agential mechanisms, not agents, a distinction I elide.

about is whether the agent can respond to the particular reason in the particular case. Fischer and Ravizza contend that an agent can respond to a reason if the agent can both recognize that reason and comport her behavior to her recognition of that reason. The moderate responsiveness conditions are evidence of those component capacities. Because Fischer and Ravizza believe that reactivity is "of a piece," they believe that weak reactivity is sufficient to show the relevant component capacity. After all, if the agent does react in some case, the agent can react in that case, and if reactivity is of a piece, then if the agent can react in that case, the agent can react in this case. By contrast, Fischer and Ravizza do not believe that receptivity is of a piece, and so the inference permitted in the case of reactivity is not licensed for receptivity. Accordingly, they offer the understandable-pattern test, which they claim is sufficient to license the inference of the relevant capacity particular to the situation. If the agent does not display an understandable pattern to his responses, then even if the agent does respond to reasons in some cases, we cannot be confident that he is doing so by virtue of the sort of appropriately functioning capacities that would license us to infer that he could respond in the particular case under consideration.

I am skeptical both that reactivity is of a piece and that an understandable pattern in other cases is sufficient to show the relevant ability in any particular case. However, I set those matters aside here. The moderateness we should seek in our threshold is not merely evidentiary. Even if it is true that the evidence of the requisite capacities is also a matter of degree, and so might also admit of a moderate

threshold, the moderate threshold I am concerned with is that of the underlying, scalar capacity.[70]

This leaves us without an analysis of the sort of moderateness that we should look for in picking out the threshold degree of reasons-responsiveness associated with binary responsibility. That should be no surprise, at least pending a better analysis of the metaphysics of the capacities at issue as well as a better analysis of the reasons for picking out a threshold at all. Accordingly, just as I appeal to the intuitive notion of a scalar capacity, I appeal to the intuitive notion of a threshold line regularly and intelligibly picked out.

2.4 The several roles for empathy

Although further details remain to be filled in, the reasons-responsiveness theory can nicely account for many of our central judgments of moral responsibility. This account is all the more promising once we notice how comfortably the reasons-responsiveness theory can accommodate our intuitions about other factors that seem to matter. Here I consider two: the role of empathy in responsible behavior and excuses grounded in irresistible urges.

---

[70] I am likewise skeptical of the other standards of moderateness offered in the literature. Brink and Nelkin, for instance, offer a moderateness standard symmetric between its cognitive and volitional elements. They look for a Goldilocks standard: "Where there is sufficient reason for the agent to act, she regularly recognizes the reason and conforms her behavior to it" (2013, p. 294). If this regularly is a matter of actual performance, then the test seems too sensitive to the agent's potentially idiosyncratic history. And if the test includes counterfactual performance, then Brink and Nelkin should explain to us what would ground that regularity, for it would seem that the regularity is grounded by the moderateness of the capacities. In any case, Brink and Nelkin might have identified a good indicator, just as Fischer and Ravizza are best understood as offering an indicator, but it is not clear that we have progressed much in explanation.

In abstract terms, empathy is "the primary means for gaining knowledge of other minds" (Stueber, 2017). When we empathize with others, we share in some of the experience they might have. We might suffer if we understand them to be suffering, and we might be pleased if we understand them to be pleased. It is controversial whether the capacity for empathy or the activity of empathy are required for moral responsibility.[71] An empathetic will is a good will, and so we might plausibly think that someone who acts without empathy thereby manifests ill will. Thus the lack of empathy or the lack of the capacity for empathy might be thought to explain blameworthiness, not to undermine it. But we should not be too quick to think that empathy cannot excuse. Accordingly, I here look at three arguments claiming that the impaired capacity to empathize can lead to compromised responsibility.

The first compromised-empathy argument claims that the proper functioning of empathy is important for the development of the reasons-

[71] Nelkin (2011) provides a helpful overview of the controversy, finally taking a position which is a significant inspiration for mine. And we see similar discussions in the debate about the relevance of psychopathy (here taking no particular stand on the relationships between psychopathy, sociopathy, and antisocial personality disorder). Though philosophers remain divided as to whether psychopathy can excuse, many see psychopathy as at least potentially grounding excuse or mitigation in large part because the psychopath lacks the capacity for empathy. For example, Fine and Kennett (2004), Levy (2007b), and Ishtiyaque Haji (2010) all claim that psychopathy interferes with moral cognition in a responsibility-compromising fashion. And Nelkin (2015) provides a helpful discussion of how thinking about psychopaths can help us clarify the relationship between attributability and accountability.

responsiveness capacities.[72] Cognitively, the exercise of empathy can help habituate us to notice morally relevant facts. Plausibly, we are drawn to notice others' interests by the operation of empathy even early in development. This might help habituate us to notice those facts. With repeated exposure, we might then come to notice moral facts more regularly and more widely, eventually noticing them without the experience of empathy. Volitionally, too, the exercise of empathy can habituate us. Early in development, the operation of empathy might lead us to act in accord with others' interests. Empathy could present those interests as mattering to us. With the repeated operation of empathy, we could thereby become habituated to take others' interests as bearing on our behavior, eventually doing so even in cases without the experience of empathy.

On this argument, empathy plays an indirect role in determining moral responsibility. What truly matters on this argument is the contemporary possession of the reasons-responsiveness capacities. A contemporary incapacity for empathy might nonetheless serve as evidence, several times removed, of some compromise in the contemporary possession of the reasons-responsiveness capacities. The contemporary incapacity for empathy might be good if defeasible evidence of a developmental incapacity for empathy, which itself would be good evidence for the lack of experienced empathy during the developmental period, which, finally, would be good if defeasible evidence of the healthy development of the reasons-responsiveness capacities. And despite the many degrees of remove between the

[72] These are speculative arguments, needing empirical work from other fields for support. For now, they are placeholders.

contemporary lack of the capacity for empathy and the contemporaneous capacities, this evidentiary chain might reasonably license a contemporary responsibility inference. Because direct evidence of the contemporary capacities is difficult to come by, and because the indirect evidence will often be mixed, the lack of a contemporary capacity for empathy might regularly be highly probative of compromised reasons-responsiveness capacities.

The second empathy argument contends that empathy plays an important helping role in the operation of the capacities. Cognitively, the experience of empathy can help to draw our attention to morally relevant facts. When, for example, we consider some self-serving action, our empathy might prompt us to attend to the harm the action might cause to others; in that way, empathy might serve to make us more reliably sensitive to at least some moral reasons.[73] Similarly, empathy might serve as a volitional aid. Sometimes it can be quite tempting to act improperly. Often it can seem that some bit of wrongdoing is quite prudential, and not all of our motivations are pure of heart. Even if we know the right way to act, sometimes it can require a strong will to act appropriately. Empathy might help in these cases. Just as the police officer's presence can help the weak-willed agent steer straight, empathy too might help keep wayward inclinations at bay. In these ways,

[73] There are sympathies between this empathy argument and the connection between the scalar view of capacities and the circumstances in which they are exercised. The capacities exist in particular contexts, and the contexts can be more or less hospitable to the operation of the capacities. The potential operation of empathy, at least in this second argument, is part of those contexts.

then, we might think that a healthy faculty for empathy makes an agent a more reliable responder to moral reasons.

Even supposing that empathy can play this sort of helping role, we need not conclude that empathy is necessary for proper behavior. That empathy can help in responding to moral reasons does not mean that lacking empathy makes it impossible to respond to moral reasons. Supposing that it remains possible for an empathetically compromised agent to respond to moral reasons, we can ask whether the lack of empathy and the resulting decrease in the capacity to respond to reasons might provide a substantial excuse even if it does not make it impossible to respond to reasons. As Brink (2013) argues, the lack of empathy could make it harder to behave appropriately without thereby making it so difficult to act rightly that excuse is warranted, so long as there are sufficient other routes to the necessary cognitive and volitional resources. For example, an agent could pick out the relevant moral reasons by looking to social cues or legal correlates, by asking for advice, or by purposefully inculcating the right habits. In order to know whether empathy's effect is strong enough to bring an agent below the moderate line, we would need to know significantly more about both how to pick out the moderate line and how empathy interacts with the operation of the reasons-responsiveness capacities. For now, it is sufficient to suspect that empathy's role might be significant enough that the lack of empathy can at least sometimes provide grounds for excuse.

The final compromised-empathy argument claims that empathy is necessary for the sort of moral understanding required for moral responsibility. This argument follows from the requirement that morally blameworthy agents have the capacity to at least genuinely understand the reasons they transgress. Mere parroting is not enough. Someone lacking empathy should nonetheless be able to anticipate the causal consequences of his actions, and someone lacking empathy should also be able to measure his behavior up against norms (though he may find it difficult to discern those norms). But this might not be enough for the sort of genuine understanding required for moral responsibility. As Neil Levy explains with respect to the empathy deficits characteristic of psychopathy, "Psychopaths know, at least typically, that their actions are widely perceived to be wrong, to be sure, but they are unable to grasp the distinctive nature and significance of their wrongness" (2007b, p. 132). Just as the experience of seeing red might seem necessary to truly understand red, the experience of empathy might seem necessary to truly understand certain sorts of moral facts.[74] An argument along these lines could establish that those who lack the capacity for empathy thereby lack the capacity to grasp the distinctive nature of moral reasons. And if they cannot grasp the distinctive nature of moral reasons, then it might also be the case that they cannot respond to those reasons as distinctively moral reasons. And if they lack the ability to respond to moral reasons *qua* moral reasons, it might seem inapt to apply the distinctive force of moral blame to their shortcomings.

---

[74] Does this require the contemporary experience of empathy or just some prior experience of empathy? I leave that question open.

That empathy can ground excuse in this way is tied to the unresolved questions about the nature of the cognitive capacity required. Recall that I distinguished between various degrees of understanding ranging from mere parroting to robust grasp of foundational grounding. Mere parroting is not sufficient, and robust understanding is too demanding. It is unclear both whether and how empathy might be relevant to genuine understanding of moral norms, and it is unclear exactly what is demanded by genuine understanding of moral norms; accordingly, resolving the question of empathy is in a large sense predicated upon resolving the more foundational question. But we might for now use our judgments about the necessity of empathy to inform our answers to the earlier question. Along these lines, Brink (2013) contends that we should hold the psychopath responsible, claiming the psychopath's limited access to the reasons to act otherwise is sufficient to give him the requisite sort of control. The psychopath can understand that his behavior is proscribed by a norm taken seriously by others, and he can understand that transgressing this norm will be taken seriously by others. If Brink is right, we should reject this empathy argument and therefore set aside the most demanding version of the cognitive requirement. However, just as I left unresolved the exact nature of the knowledge required, I leave unresolved this third empathy argument.

Thus there are at least three potential roles that empathy could play in a reasons-responsiveness scheme, none of which require augmenting the core theory of responsibility. The three roles are consistent, so accepting one does not mean abandoning any of the others. Fully defending any of these roles would require

further significant research, philosophical and empirical. In the meantime, however, the reasons-responsiveness theorist has strong grounds to claim to be able to accommodate the intuition that the capacity for empathy can matter for moral responsibility, by claiming that it matters by virtue of its interaction with the reasons-responsiveness capacities.

2.5 The excusing potential of powerful urges

Many responsibility theorists also accept that certain sorts of particularly strong desires might offer some excuse. Fischer and Ravizza offer the example of Judith, who has a "literally irresistible urge to punch her best friend, Jane" (1998, p. 231).[75] By hypothesis, the urge that leads Judith to punch Jane is not reasons-responsive: "Judith would strike Jane, no matter what kinds of reasons to refrain were present" (*ibid.*) Fischer and Ravizza conclude that Judith is not morally responsible for punching Jane. Likewise, Wallace accepts that there might be irresistible desires which compromise moral responsibility:

> [C]onsider the ... case of addiction, which is often thought to involve a susceptibility to impulses that cannot be resisted. … If these impulses are truly irresistible, then the agent will not genuinely have the ability to control his behavior in light of the moral obligations that the impulses lead him to violate. Even if he can perfectly grasp and apply the principles that support those obligations, so that he knows that what he is doing is wrong, the irresistibility deprives the agent of the capacity to act in conformity with them. … [T]o the extent that

[75] It might seem surprising that Fischer and Ravizza recognize the possibility of irresistible urges, given their claim that reactivity is "all of a piece." Remember, however, that they predicate reasons-responsiveness of particular mechanisms, not of agents on the whole. Accordingly, even if some of the agent's mechanisms are reactive, some might not be. Moreover, as I argue, an urge might be irresistible because of its interaction with the cognitive capacity. Even if the agent's volitional capacity is "all of a piece," their cognitive capacity need not be.

> irresistible impulses deprive the agent of those abilities, it would seem unreasonable to hold the agent morally accountable. (1994, p. 171)

However, while many reasons-responsiveness theorists accept that there might be excusingly irresistible urges, neither Fischer and Ravizza nor Wallace explain what it is that might make such an urge excusing.

Here, as with the role of empathy, the reasons-responsiveness account of moral responsibility has the resources to explain this intuitive feature of our blaming practices. We can begin to understand irresistible urges by returning to the two component capacities. An urge is irresistible if it mars the operation of those capacities. So, for instance, if the experience of an urge not only played the normal motivating function of desires and wants but also blinded the experiencing agent to the moral reasons to resist the urge or to mechanisms the agent might employ to resist the urge, then the urge is irresistible. The empirical psychology literature provides us with evidence that desires can have such cognitive interference effects:[76] Our attention can be driven by our desires. As Ap Dijksterhuis and Henk Aarts explain, "If one is thirsty, drinks attract more attention than things one cannot drink" (2010, p. 471). Attention is a filtering mechanism, and so when we pursue one goal, our recognition of competing goals is inhibited. This facilitates our pursuit of our goals, but it also means that our attentive and thus cognitive possibilities are limited by our desires. That a desire focuses our attention does not mean that we cannot attend to anything else. But it is plausible that some desires might so capture

[76] My discussion here largely follows Ap Dijksterhuis and Henk Aarts (2010).

attention, perhaps by virtue of their strength, that they make it significantly harder to recognize other things which matter. In the presence of such a strong desire, even the possession of an ordinarily sufficient cognitive capacity might not be sufficient to be able to recognize other reasons.[77]

Imagine a long-time smoker who might ordinarily grasp the harms of smoking when his drive for nicotine is satiated but who, in the throes of the need for nicotine, cannot be brought to think about the downsides of smoking. This smoker, at least while he is the grips of the drive to smoke, lacks the sufficient capacity to recognize the reasons not to smoke, making him non-reasons-responsive. While the literature on irresistible urges often focuses on unwilling addicts,[78] this long-time smoker might well be entitled to excuse even if he is a willing addict. After all, the long-time smoker might experience an unconflicted desire to smoke during the throes of his urge. These willing addicts are those who fail to reach the proper

---

[77] To the extent that irresistible urges compromise the relevant reasons-responsiveness component capacities by virtue of the operation of a feature common to ordinary desires, it might seem that there should be at least partial excuses in many ordinary cases. If urges excuse by way of the distraction that they cause, and if distraction is part of the ordinary and healthy operation of urges, then it might seem that all urges excuse. But that might seem too strong, since it seems implausible that so many ordinary agents are not fully responsible. I raise the worry only to set it aside. While a full defense of a reasons-responsiveness account should answer this concern, for my purposes it is enough to gesture at how an irresistible urge might excuse, raise the concern, and then appeal to the widespread acceptance of the excusing force of irresistible urges among reasons-responsiveness theorists.

[78] There are two ways that we might think an addict is unwilling. On the ordinary sense of an unwilling addict, the addict's coming to be addicted was not by willing or voluntary actions. But the literature's sense of unwilling addiction is different. Here, the addict is unwilling so long as, in the throes of the addiction, the addict does not at that point wish to be addicted.

judgment that they should resist their addiction, but they fail to reach this judgment precisely because their addiction has compromised their access to those reasons.[79]

Thinking about the volitional capacity leads to a second class of irresistible urges. If the experience of an urge not only played the normal function of desires and wants but also compromised the experiencing agent's capacity to act on contrary judgments, then the urge is irresistible. Here, the lack of significant philosophical work analyzing the volitional capacity limits how much more detailed the description can be. However, I can imagine a long-time smoker who is deeply aware of the harms of smoking and whose enjoyment of cigarettes has so long abated that he is not even distracted by the urge for nicotine. This smoker can explain to you why he should not smoke and even that he does not want to smoke, and he can explain all of this while he is smoking. However, this smoker smokes as if an automaton. While in the throes of the urge, he lacks the capacity to act on his normative judgments, making him non-reasons-responsive. The irresistible urge which compromises the volitional capacity thus makes ready sense of the unwilling addict.[80]

---

[79] For some, the relevant sorts of irresistible urges must meet two conditions: a) they must be in some sense irresistible, and b) they must conflict with some important aspect of the agent's practical psychology. For example, John Christman (1991, p. 16) writes, "What is problematic about compulsive desires is not merely that they are compulsive--uncontrollable at the moment of effectiveness--but rather that they often are in manifest conflict with the agent's other desires." Christman is concerned with the conditions of autonomy, but insofar as I am concerned with moral responsibility, I reject that second sort of condition.

[80] To be more precise, the compromise of this volitional ability excuses the addict from his behavior in the throes of the addiction. The unwilling addict can still be blameworthy if the addict is responsible for wrongfully becoming or allowing

And we should not be distracted by the vocabulary which populates this debate. An urge need not be literally beyond all possible resistance to ground excuse. An agent might still retain some capacity and yet be below the level of the moderate capacity. Consider again the addict. We might suspect that all but the very worst of the addicts could resist their addictions if the countervailing reasons were pressing enough. The addict might not heed his own welfare, but this does not mean that there might not be more pressing reasons which could change his behavior, such as the presence of law enforcement. That the addict retains some degree of control, however, does not mean that the addict retains moderate control. Moreover, the excuses which can be grounded by these urges come in degrees. The worst addictions might excuse completely, but many strong addictions would excuse partially.

While most urges that are irresistible are likely to be strong urges, not all irresistible urges need be strong urges. What matters is whether the urge interferes with the agent's reasons-responsiveness capacities, and so there might be strong urges which are resistible, and there might be moderate or even weak urges which are irresistible. We should be careful not to extend an excuse just because an urge is a very strong. That an urge is very strong might make it almost certain that the agent will act upon the urge, but this is in large part because very strong urges are

---

himself to remain addicted. In that case, we could see the unwilling addict's behavior in the course of his addiction not as the wrongdoing for which he is originally responsible but as the consequences of the prior wrongful behavior for which we may hold him responsible. Wallace advances an argument along these lines looking primarily at the willing addict, though the argument applies equally to both.

likely to outweigh other urges, even other appropriately recognized urges. This does not make the urge irresistible. Instead, we should excuse only if we have reason to think that the strong urge, in addition to (and perhaps because of) being strong, also has a compromising effect on the agent's reasons-responsiveness capacities. Likewise, some moderate or mild urges might, despite seemingly readily superable to outsiders, interfere with an agent's reasons-responsiveness capacities. This might be especially true, for instance, of the urges associated with certain sorts of habits. Accordingly, while strength might be a fair predictor of irresistibility, it is not what matters.

Finally, it is important to recognize that there is more than one way of resisting an urge.[81] An agent might lack the capacity to resist in one way but possess the capacity to resist in another way. That undermines the agent's claim to excuse. When confronted with some desire, we might act otherwise either by acting upon some other desire (thereby circumventing the original desire) or by intentionally resisting the desire. And when we act intentionally to resist a desire, we have options. We might conquer it by brute force, willing ourselves to resist the desire, or we might conquer it by skilled strategy, intentionally distracting ourselves, focusing elsewhere, and the like. Thus, as Al Mele (1990) has argued, an irresistible urge is one which makes all of a number of alternative actions impossible. Accordingly, an agent might lack the capacity for resisting an urge by brute force and yet be responsible for acting on the urge, so long as the agent possessed the needed

[81] The best treatment, which I largely follow, is Al Mele (1990).

capacities to resist the urge in some other fashion. The cigarette smoker who knows that he cannot force himself to refrain from smoking once he has the cigarette to hand might nonetheless be responsible if he knows that he could distract himself by running or the like but fails to pursue that alternative strategy.

2.6 Reasons-responsiveness and control

The reasons-responsiveness account of moral responsibility thus laid out offers an appealing explanation of our patterns of blaming and excusing. The account remains so far a promissory note, wanting for fuller development of the notion of capacity at its heart as well as more sophisticated sociological and philosophical work connecting the reasons-responsiveness capacities to our practices. However, especially when we attend to the familiar component capacities and when we see those capacities as scalar, the account offers a plausible and attractive explanation of our practices, and it gives us promising programs for further investigations of the conditions of responsibility, for diagnosing particular agents, and for determining exactly when particular conditions should or should not excuse.

Moreover, the reasons-responsiveness account of moral responsibility offers an attractive sense of the sort of control intuitively important for moral responsibility. That appropriate praise and blame are tied to control has long been recognized. As Aristotle explained, it is "on voluntary passions and actions [that] praise and blame are bestowed" (1998, 1109b). This is quite close to the reactive attitudes account of moral responsibility, especially if we think voluntary actions are

those performed while in possession of the capacities for voluntary control. Aristotle gives us the example of someone carried away by the wind. If you later blame the agent carried away by the wind for having left, you can imagine the agent defending himself by explaining that he was forced to leave. He might say, "Because I was forced by the wind, and because I had no control over the wind, I had no control over not being where you expected me to be. You should not blame me." Thus, control seems to be an important condition of moral responsibility.

Contemporary philosophers have sometimes seen in the appeal to control a role for the Principle of Alternative Possibilities. On that principle, an agent has control over an action in the sense necessary for blame only if there are alternative courses of action open to the agent.[82] If not, then the agent did not have control, and the agent should not be blamed. The agent carried by the wind, for instance, did not have any other significant possibilities available to him. No matter what he did, the wind would have carried him away.

But the could-have-done-otherwise principle seems to run afoul of Frankfurt cases, cases where an agent could not have done otherwise only because of a counterfactual intervener who in fact contributes nothing other than ensuring that the agent could not have done otherwise. In Harry Frankfurt's words:

> Suppose someone--Black, let us say--wants Jones to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until Jones is about to make up his mind what to do, and he does nothing unless it is clear to him (Black is an excellent judge of such

[82] For more on the Principle of Alternative Possibilities, see Peter van Inwagen (1975) and Harry Frankfurt (1969), among many others.

> things) that Jones is going to decide to do something *other* than what he wants him to do. If it does become clear that Jones is going to decide to do something else, Black takes effective steps to ensure that Jones decides to do, and that he does do, what he wants him to do. … Now suppose that Black never has to show his hand because Jones, for reasons of his own, decides to perform and does perform the very action Black wants him to perform. (1969, pp. 162–163 subscripts omitted)

There has been a mammoth literature refining, critiquing, and exploring the Frankfurt cases. But many, Frankfurt and myself included, think it intuitive that Jones is just as responsible for his action as he would have been had Black been absent. Jones seems responsible even though Black's presence makes it the case that Jones could not have done otherwise.

Aristotle's conditions of the voluntary and the involuntary can explain Jones's responsibility. Recall that Aristotle offered two formulations of the conditions of involuntary action: an action is involuntary if it is forced by outside conditions or if it is done without any contribution from the agent. That Jones performed the action on the basis of his own reasons-sensitive decision means that Jones made a true contribution. And while Black remained ready to compel Jones's performance, Black did not in fact compel Jones's performance, and so we should accept our intuition that Jones's action was not forced. As Fischer and Ravizza explain, we should see the Frankfurt cases as distinguishing two notions of control, guidance control and regulative control (1998, pp. 31–34). Frankfurt agents like Jones lack regulative control, because they cannot bring about any alternative results, but they possess guidance control, for they make a contribution to the action's coming about. The

intuition that Frankfurt agents like Jones are responsible shows us that responsibility is about guidance control, not regulative control.

The reasons-responsiveness capacities provide a satisfying explanation of the guidance control which marks Jones's contribution to his action.[83] Jones has control over his action because his action stems from his recognition of certain reasons that bear on his choice and his volition based upon that recognition. Moreover, we are given no reason to think that Jones lacks the relevant component capacities to recognize or react to the ordinary reasons not to kill. That Black waits in the wings to meddle, should Jones fail to recognize the reasons that lead him to act as he does, fail to care about those reasons, or fail to act upon those reasons, guarantees that Jones does not act otherwise, but it does not mean that Jones fails in any of those ways. If Jones lacked the capacity to recognize or react to those reasons, he might still act as he does--after all, Black might so compel him. The possession of those capacities is necessary for Jones's behavior to be the product of Jones's agency in the right sort of way, and so the presence of those capacities can provide a satisfying explanation of the intuitive pull of the Frankfurt cases, and thus of the conditions of moral responsibility.

These capacities can even serve to provide a satisfying, compatibilist reading of the Principle of Alternative Possibilities.[84] The Principle of Alternative Possibilities requires that other possibilities be open and available. But what is required for a possibility to be open and available is up for debate. The

---

[83] This is Fischer and Ravizza's central and important contribution.
[84] I thank David Brink for pushing me to think more about this.

incompatibilist wants to claim that a possibility being open and available means that it is not foreclosed by determinism, the laws of nature, the past, and the rest. But the reasons-responsiveness theorist might claim instead that a possibility is open and available to an agent when some similarly situated agent with the same capacities and in the same circumstances could bring about the possibility. If so, then some deterministically unavailable possibilities would nonetheless be open and available. That is, there is a distinctively reasons-responsiveness way to read the Principle of Alternative Possibilities.

That the reasons-responsiveness account can provide a satisfying explanation of the Principle of Alternative Possibilities reflects the broader fact that the reasons-responsiveness account can provide a satisfying sense of the sort of control intuitively involved in our ascriptions of responsibility. That explanatory success plus the extensional promise of the account in picking out the paradigmatic cases of responsibility, compromised responsibility, and irresponsibility render the reasons-responsiveness account of moral responsibility compelling.

Chapter 3 The question of history

3.1 The disputed role for history

The possession (or not) of the reasons-responsiveness capacities is not a fixed fact. It is an ordinary fact that an agent who lacks the capacities as a child can later possess them as an adult, some mental illnesses might strip away the capacities from an agent who once possessed them, and some mental illnesses might be treated. The capacities are possessed (or not) by an individual at some particular time. This means that we need to know which time or times matter in order to evaluate an agent's responsibility for some behavior. As a starting point, we might think that the relevant time is the time of the action. We can see this in the general contour of the excuses we entertain. Being incapacitated prior to the behavior is ordinarily no excuse at all, barring that the incapacitation had some effect which continued through the time of the behavior. We blame adults after all! Likewise, while being incapacitated after the behavior at issue raises problems about the expression of blame, once the wrongdoer returns to competence, we ordinarily resume the expression of blame. Thus, as a starting point, it is the agent's capacities at the time of the behavior which matter.

But our intuitions and patterns of response also suggest that an agent's history can matter for his responsibility. There are two sorts of cases which particularly support this suggestion: tracing cases and bad-history cases. In the tracing cases, an agent lacks the requisite reasons-responsiveness capacities when he acts, but he is responsible for lacking those capacities. The standard example is

the extremely intoxicated wrongdoer. This wrongdoer's intoxication mars his normative psychology, undermining the component capacities which make for reasons-responsiveness. If reasons-responsiveness at the time of the behavior is a necessary condition of moral responsibility, then it seems that this agent is not responsible. However, many severely intoxicated wrongdoers are intuitively blameworthy, and so it appears that, whatever role reasons-responsiveness at the time of the behavior might play in a theory of moral responsibility, contemporary reasons-responsiveness may not be a necessary condition of moral responsibility.

If the tracing cases challenge the thought that contemporary reasons-responsiveness is necessary for moral responsibility, the bad-history cases challenge the thought that contemporary reasons-responsiveness is sufficient for moral responsibility. One particularly powerful example of this is the wrongdoer who had been abused and neglected as a child. This background seems to mitigate or block the agent's blameworthiness even if the agent developed the reasons-responsiveness capacities. Our intuitions about these bad-history cases thus suggest that history can tell against responsibility, just as our intuitions about the tracing cases suggest that history can tell for responsibility.

The implications of incorporating history go beyond accounting for these central cases. Tracing is widely seen as an anodyne if necessary element of any satisfying theory of moral responsibility, one that may be incorporated into most theories of moral responsibility without risking further complications. The bad-history cases, however, are often seen as the wedge by which a powerful argument

for incompatibilism can be introduced. This is most plain in Pereboom (2001, 2014). Such incompatibilists argue that the best explanation for the intuition that a bad-history agent is not culpable is that her wrongdoing is the product of matters outside of her control. If it is plausible that the bad history is the source of the agent's wrongdoing, and given that it seems clear that the bad history is, ordinarily, outside of the agent's control, then we might plausibly advert to the lack of control over a determining factor to explain the lack of responsibility. That lack of control is particularly noticeable in the bad-history cases, but it is plausibly true of all of us that our behavior is the product of factors over which we lack control. If so, goes the argument, none of us is responsible.[85]

So we should ask whether history matters. Fischer and Ravizza (1998, p. 170) see this question as contrasting two classes of theories: historical theories and

---

[85] Here, I have presented the argument to incompatibilism as:

1. The bad-history agent is not responsible.
2. The best explanation of the bad-history agent's not being responsible is that his behavior is the product of factors beyond his control.
3. That explanation extends to all of us.

C. So no one is morally responsible.

Some incompatibilists present the argument to incompatibilism as:

1. The bad-history agent is not responsible.
2. There is no plausibly relevant difference between the bad-history agent and the rest of us.

C. No one is morally responsible.

I elide the differences between these two arguments.

time-slice theories.[86] Fischer (2000) offers examples contrasting the two classes in other domains: size, weight, height, shape, and symmetry are intuitively time-slice phenomena, while sunburns and counterfeiting are intuitively historical phenomena. Frankfurt (1971) presents perhaps the most famous ahistorical view of moral responsibility in his mesh account. For Frankfurt, an agent is responsible so long as there is a sufficient mesh between the agent's first-order desires, his second-order desires, and his actions. It does not matter how those desires and that mesh came about:

> The causes to which we are subject may also change us radically, without thereby bringing it about that we are not morally responsible agents. It is irrelevant whether those causes are operating by virtue of the natural forces that shape our environment or whether they operate through the deliberative manipulative designs of other human agents. We are the sorts of persons we are; and it is what we are, rather than the history of our development, that counts. The fact that someone is a pig warrants treating him like a pig, unless there is reason to believe that in some important way he is a pig against his will and is not acting as he would really prefer to act. (Frankfurt, 2002, p. 28)

Frankfurt, a mesh theorist, takes a hard ahistoricist line. But we should ask whether a similar line is attractive for a reasons-responsiveness theorist.

---

[86] Probably in part because different theorists have different notions of the sort of history which might be philosophically interesting, the vocabulary in this debate is varied. The historical theories are sometimes also called external theories; the ahistorical theories are sometimes also called internal theories, structuralist theories, and time-slice theories.

I begin by distinguishing between ways that history can matter. No one should deny that there are some inoffensive ways history could matter.[87] History certainly plays important causal roles. As Fischer and Ravizza describe it, history casts a shadow (1998, p. 194). That some agent is reasons-responsive now is the product of the agent's past--the way the agent was raised, the properties the agent inherited at birth, etc. That some agent has a certain quality of will now is likewise the product of the agent's past--the way the agent was raised, the properties the agent inherited at birth, etc. Even if these history facts are not wholly determinative of the agent's possession of the capacities and quality of will now, any plausible theory of moral responsibility concerned with either the reasons-responsiveness capacities or Strawsonian quality of will must admit that history facts are causally relevant.

Historical facts are also diagnostically and epistemically relevant. If we want to know whether some agent possesses the reasons-responsiveness capacities at some particular point, we might profitably look to the agent's history. If, as is more than plausible, certain pasts tend to yield healthy reasons-responsiveness capacities more than others, we can use evidence regarding an agent's past as evidence of the present possession of healthy capacities. And it is also more than plausible that the capacities are relatively stable. If that's so, we might use evidence of the agent's possession of the capacities at one time as evidence that the agent possessed the

[87] Fischer and Ravizza (1998, pp. 170–194) give an especially clear account of the different roles history might play, and what follows borrows liberally but not exactly from them.

capacities at some other time--with the particular weight of the evidence varying based upon factors like the separation of the two times, intervening events, and the like. For both of these reasons, the agent's history would be important diagnostic evidence of the agent's moral responsibility, because the agent's history would be good evidence of the agent's possession of the reasons-responsiveness capacities at the time of the behavior.

There are other historicist roles I'd like to mark and set aside, roles suggested by the time-slice contrast. First, the time-slice contrast suggests that history is only one of two non-time-slice elements: there is also the future.[88] And so we could ask whether what happens after a bit of behavior affects the agent's responsibility for that behavior. Plausibly, what follows a particular bit of behavior does affect whether and how we should respond to the agent. We widely treat consequences as relevant to our responses. The criminal law treats murderers more harshly than attempted murderers, and the merely reckless driver is castigated less severely than the reckless driver who causes a serious accident. This feature of the criminal law is consistent with how we treat each other in our interpersonal relationships. I resent someone who actually harms me more than I resent someone who intends to harm me but never gets the chance. Consequences are not the only way that events subsequent to the initial behavior affect whether and how we respond. How the agent and those impacted by the agent act with respect to the

[88] The future is, of course, not history. However, both the future and the past are readily contrasted with the time-slice, and so asking whether responsibility is a time-slice phenomenon entails asking about both the past and the future.

agent's wrongdoing often affects how we hold him to account for that behavior. When an agent repudiates, apologizes, or makes amends for a bit of wrongdoing, we tend to blame him with less force, if at all. Similarly, if those harmed by the agent forgive the agent, we tend to blame him with less force, if at all.[89] And we might also look to the future to determine whether the agent who acted wrongly is the same agent we might now blame. Benjamin Matheson (2014) raises this question, claiming that there are serious questions about personal identity which should be addressed in determining whether to blame.

I set aside questions about the relevance of the future to responsibility. While philosophers like Fischer and Ravizza often treat (perhaps only casually) time-slice accounts as the relevant contrast to history being relevant, rarely is the future considered alongside the past. And we can make sense of the distinctive treatment of the past. The past seems to constrain our control in a way that the future cannot. Insofar as control has seemed to matter tremendously for most theorists about responsibility, it is no surprise that the past has interested many theorists about responsibility. Accordingly, I too set aside questions about the future, albeit leaving them for future investigations, not rejecting them as irrelevant.

There are two other ways that the time-slice contrast might lead us astray. First, a particularly thin notion of a time slice might be insufficient to capture all of the facts which would pick out the relevant actions, intentions, and the like.

---

[89] Recent work on these phenomena--and especially forgiveness--has been particularly significant in my own thinking about blame. For example, my appreciation of the seeing-as element of the reactive attitudes was largely sparked by thinking of how forgiveness relates to blame, as urged by Allais (2008a, 2008b).

Plausibly, no actions are of an instant. Even a quotidian battery, the ordinary punch, takes some time to occur. Similarly, a particularly thin notion of a time slice might exclude the ordinary interaction between intention and activity, surmising that intentions play a causal role and surmising that causal interactions take place over time. As David Zimmerman explains, "reasons-responsive decision-making is a *process* which takes up time and is thus not a 'current time-slice' phenomenon" (2002, p. 210, emphasis in original).[90] Accordingly, we might broaden the time-slice view into a 'chunky' time-slice view, wide enough to incorporate whatever is minimally necessary to pick out the relevant actions, intentions, and the like. Doing this would not seem to incidentally incorporate the sorts of history involved in the tracing and bad-history cases.

Then, even if we widen the time-slice view enough to accommodate the full action, we might not include enough time to be certain that it is an agent which we are holding responsible. Plausibly, to be an agent requires persistence.[91] Perhaps there are agential features, like elements of personality, which must endure stably (even if not perfectly) for there to be an agent at all, or perhaps being an agent requires causal connections between how one was and how one is. A wholly unstable being might be no agent at all, but instead a series of instant quasi-agents

---

[90] Fischer (2004) himself later divided theories of responsibility into time-slice, interval, and deeply historical theories. Recognizing the role of intervals of history is how I understand Vargas's comment that "Some structural features of agency (deliberation, for one) have a certain amount of temporal extendedness" (2006, p. 355).

[91] This seems to be the case to me, though the literature on the question of history includes a number of instant agents whose plausibility might seem to challenge the supposition.

or a single, radically unstable quasi-agent. It is plausible that the persistence requirements of agency are more demanding than the historical requirements of actions, intentions, and the like. If so, then a case where the facts were sufficient to ground the reasons-responsiveness capacities as well as an action, an intention, and the like might all be present, but because of history even further back, there might be no agent to hold responsible.[92] This can make for a further, deeper role for history. But here, as with the broadening potentially required to account for actions, intentions, and the like, we might allow that history matters for establishing the existence of the relevant agent without thereby implicating the sorts of history involved in the tracing and bad-history cases.

While the notion of a time-slice account of what matters for moral responsibility might provide a good-enough rough contrast to the accounts of moral responsibility for which history plays an essential role, there are too many ways history might plausibly matter for the time-slice notion to be any more than that rough contrast. Instead, the question of history I am interested in is picked out by normative concerns: getting blame right, addressing concerns about determinism, and identifying the right explanation of the importance of the reasons-responsiveness capacity. As Ishtiyaque Haji (a historicist) writes, we are asking whether "moral responsibility's key conditions can[] be specified independently of facts about how the person acquired her responsibility-grounding [or voiding]

[92] Of course, it might be that a certain phenomenon only counts as, e.g., an intention if it is predicated of an agent. In such a case, the history required for there be an intention would be at least as broad as the history required for there to be an agent.

psychological elements" (2013, pp. 185–186). With that in mind, the argument should revolve around the two motivating sorts of cases, the tracing case and the bad-history case. Instances of both cases seem to generate intuitions which gainsay the ahistoricist, reasons-responsiveness account of moral responsibility. Accordingly, the argument for historicism will be augmented if I show that our intuitions in those cases can be accommodated. To this end, I will show that those intuitions can comfortably be explained as the product of the shadows history casts upon the present, as marking something other than responsibility and blameworthiness or praiseworthiness, or as reasonably expectable errors.

3.2 Fischer and Ravizza's taking-responsibility account

Before turning to that task, I want to examine the taking-responsibility element of Fischer and Ravizza's account of moral responsibility. The taking-responsibility account is interesting in its own right, but I consider it here also to show how a theory of moral responsibility might accommodate the interests that often motivate historicism (intuitions about certain kinds of cases but also other concerns) without abandoning the ahistoricism of its central reasons-responsiveness commitments.

For Fischer and Ravizza, reasons-responsiveness contemporaneous with the relevant action is neither necessary nor sufficient for moral responsibility. Reasons-responsiveness is not necessary for moral responsibility for Fischer and Ravizza because they accept tracing (which I take up in the next chapter). Reasons-responsiveness is not sufficient for moral responsibility for Fischer and Ravizza

because they contend that an agent must have previously taken responsibility for the relevant elements of her moral psychology. Fischer and Ravizza argue that we cannot be held responsible for our behavior unless we have come to have two important self-directed attitudes. To have the first attitude, which I will call the causal attitude, the agent must take herself[93] to be causally efficacious. To have the second attitude, which I will call the target attitude, the agent must take herself to be a proper target for judgments of moral responsibility. For Fischer and Ravizza, the mere possession of these two attitudes is not sufficient--the agent must have come to have the attitudes in the right sort of way; this is part of the historicism in their account of moral responsibility.

Begin with Fischer and Ravizza's second attitude, the target attitude. An agent has the target attitude when the agent "accept[s] that he is a fair target of the reactive attitudes" (1998, p. 211). Fischer and Ravizza argue that we might expect that an agent who possesses the target attitude will respond to blame by offering a justifying or excusing explanation, by experiencing guilt, and the like.[94] An agent who lacks the target attitude, however, might fail to take up the import of the reaction, ignoring, rebuking, or becoming confused by the expression of praise or blame. In such cases, the agent has no sense that blame or praise could be appropriate for them, and so Fischer and Ravizza say that sanction will be ineffective or inappropriate in the absence of this belief.

---

[93] Here, again, I am eliding distinctions that occupy Fischer and Ravizza in part because of their particular concern with mechanisms rather than agents.

[94] That is, an agent who possesses the target attitude is expected to respond to blame in the dialectical ways explored by McKenna (2012).

I deny that the possession of the target attitude is a condition of moral responsibility. I grant that the reactive attitudes depend upon target uptake to function normally, but this does not mean we should complicate our conditions of responsibility. Because consideration of the target attitude should not lead us to add any complications to our theory of responsibility, consideration of the target attitude should not lead us to add any historicist complications. First, as I laid out in Chapter One, the reactive attitudes have roles other than communication to the actor. Even if that communication role is frustrated in some particular case, the attitudes might yet fulfill their other roles. Second, and relatedly, this concern confuses matters of internal and external justification. That the reactive attitudes might fail in some particular instance to fulfill the roles that can support the external justification of the practice does not entail that the reactive attitudes are not internally justified in that instance. Third, this concern has to do with the appropriateness of expressing the attitudes, but the appropriateness of expressing the attitudes is not the same as the responsibility of the agent. This is an apparent instance of the commission of the moralistic fallacy, confusing appropriateness with fittingness. We should take seriously the receptive characteristics of the agents we blame, since uptake is important. But this does not have to do with responsibility itself.[95]

---

[95] David Brink has pointed out to me another potential role for the target attitude, a causal role. It might be that, in order to develop reasons-responsiveness, the agent must have the target attitude. That the target attitude might play a causal role along these lines is plausible (although not obvious). However, even supposing it is true that the target attitude does play such a causal role, we do not need to complicate

The causal attitude, by contrast, seems to have more to do with responsibility. To have the causal attitude picked out by Fischer and Ravizza, the agent "must see himself as the source of his behavior…; he must see that his choices and actions are efficacious in the world" (1998, p. 210). An individual who had no sense that his choices could have effects in the world would therefore lack the proper understanding of the import of his choices required for moral responsibility. What would someone look like who lacked the causal attitude? Fischer and Ravizza appeal to Daniel Dennett's example of someone who, having jumped off of the Golden Gate Bridge, wonders whether having done so was a good idea (Dennett, 1984, p. 104).[96] The jumper's deliberations can (presumably) have no effect at this point--and so long as the jumper recognizes this, that recognition is the abandonment of the causal attitude.[97] According to Fischer and Ravizza, an agent who lacks the causal attitude is "essentially passive, buffeted by forces that assail him" (1998, p. 221). They offer the example of a sailor in a storm who does not believe his rudder is functioning; because he does not think he can have any effect, he just allows the winds to drive him on. Although an agent who globally lacked the

---

the theory of responsible. That would be an instance of shadow-casting, of the past being relevant to the present because of the effects of the past upon the present.

[96] I register my discomfort with the casual treatment of suicide and with labeling the case of a suicide a case of "local fatalism," in Dennett's words. This sort of case--where action set in motion can no longer be retracted--is quite ordinary, and casual discussion of the man's "future destination" (1984, p. 104) is untoward.

[97] It is important to distinguish here between the fact that the jumper's deliberations are causally inefficacious and the jumper's recognition of that fact. Both matters are plausibly relevant to Fischer and Ravizza's scheme, but here they are concerned with the recognition.

causal attitude would scarcely be an agent, all of us are regularly at least locally like the sailor, lacking the causal attitude.[98]

But accepting that the causal attitude is important does not entail accepting a historicist account of moral responsibility. The mere possession of an attitude does not seem to require any more history than the mere possession of the other phenomena which make up an action, such as intentions. So we could accept that the causal attitude is required but just see that as another element needed in the time slice.

The historicism of Fischer and Ravizza's taking-responsibility account is in the third element, the requirement that the agent come to have the two requisite attitudes in the right sort of way. I've argued that the time-slice versions of the attitudes can serve the functional roles Fischer and Ravizza identify. Nonetheless, Fischer and Ravizza argue that the beliefs must be "based on … evidence in an appropriate way" (1998, p. 236). They describe "the long, complex, and difficult process of moral education" (1998, p. 208, quotation marks omitted). This involves things like reacting to children with faux reactive attitudes, treating them as if they are responsible agents in order to teach them what it is to be a responsible person

[98] That locally lacking the causal attitude is so widespread raises questions about how we should treat people who are themselves responsible for lacking the causal attitude. How should we regard, for instance, people who falsely think to themselves, "What a horrible situation, and it's a shame there's nothing that I can do"--especially if we think that they should, with even only slight reflection, be able to see that there is much that they could do? One tempting solution might be to impute the causal attitude to such agents so that we can hold them responsible. But, to anticipate my later argument, I reject this sort of tracing strategy elsewhere, and so I reject it here as well.

and to accustom them to thinking in such terms. As Fischer explains, "a child, fairly early on, realizes that when he chooses to punch his sister and he moves his arm in such a way that his sister is hit, she cries as a result of his choices and bodily movements" (2000, p. 389). At the same time, we invite children to see that our reactions (including the faux reactions) are a function in part of the manifestation of their willpower, of their choices and actions. In this way, the child comes to see himself as a moral agent in the way putatively required for holding the child responsible.

Fischer and Ravizza's brief account of the nature and role of the moral education of children is not implausible, even if best seen as a placeholder for a substantive study. It would not be surprising if our self-regarding beliefs were largely a function of how we are raised, of how adults both purposefully and incidentally induce us to regard ourselves when we are young; in fact, it would be surprising were this not the case. Indeed, it is likely that certain sorts of upbringing are all but necessary for an agent to have these attitudes. Late-in-life inculcation of these attitudes may be difficult or even psychologically impossible. Therefore, I readily accept that the past is causally relevant to the contemporary possession of Fischer and Ravizza's two agential attitudes.[99]

---

[99] I do not mean to suggest that the taking-responsibility scheme is uncontroversial other than its claim to essential historicism. In particular, the scheme is tied closely to Fischer and Ravizza's arguments about mechanisms and mechanism individuation, and those are controversial claims. See, for instance, Seth Shabo (2005). And there are worries about whether an agent could intentionally avoid culpability by abandoning either of the beliefs. See, for instance, Neal Judisch (2005).

Fischer and Ravizza see a thicker role for their third requirement than a causal or evidential role. They appeal to reliabilist standards of epistemic justification, insisting that the third element is an important, independent element.[100] But the dialectical justification of the third requirement is the need to address bad-history cases.[101] Accordingly, given that we can satisfy all of the functional desiderata of the taking-responsibility account without introducing a historicist element to the theory of responsibility, we should admit this particular historicist element only if we lack alternative, ahistorical explanations of the cases which challenge historicism. I now turn to that challenge.

---

[100] As Haji (2000) points out, there are suitable internalist epistemic standards. See David Zimmerman (2002) for similar concerns.

[101] Fischer (2004, 2014) argues that we need the third element to avoid manipulation of the two attitudes. See Eleanor Stump (2002) for a statement of the sort of manipulation-of-taking-responsibility concern Fischer is addressing. The manipulation of taking responsibility might seem more worrisome than the manipulation of the other components of an agent's practical psychology. If these two attitudes escape manipulation, then even if other features of the agent's practical psychology are the product of manipulation, the agent would have appropriately considered the features and incorporated them into her psychology. It's not clear that this is enough to avoid manipulation concerns, but since I think those concerns are adequately addressed without the complication of the taking-responsibility scheme, I set aside this concern.

Chapter 4 Resisting tracing's siren song

It is an unfortunately familiar fact that intoxication can lead people to act improperly, even criminally. Consider Elie Joseph Arsenault.[102] In June 1954, Arsenault shot and killed Harriet Hinckley, his drinking partner and intimate. The murder occurred at the end of a days-long drinking binge, after Arsenault and Hinckley had shared 11 pints of liquor accompanied by barbiturates. Arsenault remembered little--not shooting Hinckley, not telephoning the police, not confessing to the crime.

Incapacitated wrongdoers like Arsenault pose a problem for reasons-responsiveness accounts of moral responsibility. Those accounts are powerful and popular--it seems right that moral responsibility depends upon an agent's having the capacities to perceive and act upon moral reasons. But, while Arsenault's intoxication likely incapacitated him at the time of his crime, surely he should not be excused from blame. In order to address such cases, many reasons-responsiveness advocates include a tracing condition to supplement the ordinary conditions of responsibility. The intoxicated wrongdoer is blameworthy despite his incapacitation precisely because he is responsible for becoming incapacitated. We hold him responsible for his intoxicated wrongdoing by tracing back to his responsibility for becoming intoxicated. Arsenault was responsible for becoming intoxicated, and his responsibility for the later murder of Hinckley can be traced back to that earlier responsibility.

---

[102] *State v. Arsenault*, 152 Me. 121, 124 A.2d 741 (Maine 1956).

But not everyone has accepted tracing. One challenge notes that we do not need tracing to blame culpably incapacitated agents. We can hold these agents accountable for their behavior in becoming incapacitated and for the foreseeable consequences of that behavior, and we can do so without taking on the apparent costs of tracing. I go further--I claim that tracing gets things wrong. To show this, I consider a different sort of case: the Odysseus case. Odysseus incapacitated himself in order to sail safely past the Sirens. Arsenault's incapacity was improper; Odysseus' was not. But things might have worked out poorly for Odysseus; he might have committed some wrongdoing while incapacitated. If so, then my intuition is that Odysseus would have been unlucky but not blameworthy. The core reasons-responsiveness account agrees, but tracing accounts expose unlucky Odysseus agents to blame. Since reasons-responsiveness responsibility appears to get us what we want (as urged by the first challenge) and tracing gets us verdicts that we do not want (as shown by my new challenge), we should reject tracing.

4.1 Ordinary responsibility and the motivation for tracing

Tracing is often offered as a supplement to reasons-responsiveness accounts of moral responsibility. Many reasons-responsiveness theorists deny that moral responsibility can be an ahistorical evaluation because they see tracing as a necessary element of any plausible account of moral responsibility. Fischer and Ravizza, for instance, consider Max, a drunk driver. Max drinks so much that he is "almost oblivious to his surroundings" (1998, p. 49). Intoxicated, he attempts to drive home and, unfortunately, strikes and kills a child in a crosswalk. By

assumption, Max's intoxication left him non-reasons-responsive, both when he decided to drive and when he struck and killed the child. This suggests that Max might be excused under the ordinary-responsibility account's ahistorical analysis of responsibility. Nonetheless, intuitively, he is to blame. Cases like Max's suggest that ordinary responsibility is explanatorily inadequate.

Fischer and Ravizza explain that drunk drivers like Max are responsible for their intoxicated behavior because they are responsible for becoming intoxicated. Fischer and Ravizza hold culpably incapacitated agents responsible for their culpably incapacitated wrongdoing by tracing responsibility for the wrongdoing back to responsibility for the prior behavior that led to the incapacity. As they explain:

> When one acts from a reasons-responsive mechanism at time *T1*, and one can reasonably be expected to know that so acting will (or may) lead to acting from an unresponsive mechanism at some later time *T2*, one can be held responsible for so acting at *T2*. (1998, p. 50)

Tracing allows us to hold an incapacitated agent responsible for her wrongdoing so long as there was some prior moment when the agent could act to avoid her incapacitation and could reasonably foresee her subsequent incapacity and wrongdoing. I will call the account of reasons-responsiveness responsibility supplemented by tracing "the tracing account," noting that the tracing account includes both traced responsibility and ordinary responsibility.

The tracing account offers to address the explanatory inadequacy that appeared to threaten ordinary responsibility. Return to Max, Fischer and Ravizza's drunk driver. By hypothesis, Max was not incapacitated at the time he was drinking,

and he could reasonably have been expected to know that drinking to excess could lead him to act wrongly while incapacitated.[103] When he did later drive while intoxicated to the point of incapacity, we ground responsibility for the intoxicated driving by tracing back to his responsibility for drinking to the point of intoxication.

In addition to appearing to explain our intuitions about culpable-incapacity cases, tracing tracks the way courts have regarded intoxication as a defense.[104] The Model Penal Code, prepared as an advisory guide for legislatures and courts, provides an affirmative defense to defendants whose intoxication interferes with their cognitive or volitional normative capacities, but it allows this defense only when the intoxication was not self-induced.[105] Applying similar reasoning, the Supreme Court of Minnesota addressed the responsibility of a hit-and-run driver

---

[103] In Fischer and Ravizza's mechanism-specific terminology, we would say that Max's drinking was the product of a reasons-responsive psychological mechanism.

[104] The claim that moral responsibility and criminal responsibility are related is a substantive and potentially controversial but widely accepted claim. I accept but do not here defend the view that moral responsibility is a normative condition on the appropriate finding of criminal responsibility. Moreover, as George Vuoso writes, "Whatever one's position on whether moral and legal responsibility are logically related, it is a plain fact that in practice our criminal law is such that people are generally held criminally responsible only when they would also be held morally responsible" (1987, p. 1663).

[105] Section 2.08(4). We should distinguish two sorts of defenses available at law: elemental defenses and affirmative defenses. An elemental defense contends that one of the crime's constitutive elements is missing; it denies that the defendant has committed the crime. An affirmative defense, by contrast, does not dispute that the crime was committed, but it denies that the defendant should be held accountable for the crime. It is not controversial that intoxication, self-induced or otherwise, can ground an elemental defense. Consider burglary: A defendant commits burglary by breaking into a building with the intent to commit some further crime. If a defendant is so intoxicated that he cannot form the intent to commit some further crime, then no burglary has been committed (though the defendant may be guilty of trespass). However, the tracing question arises when we consider the conditions of responsibility, and denials of responsibility constitute affirmative defenses.

who claimed to have been unwittingly incapacitated by his prescribed medicine.[106] The court ruled that the medicated defendant could be excused from responsibility if two conditions had been met: (a) the medication caused him to be temporarily insane and (b) he neither had known nor had good reason to know that it would have this effect. If, that is, the medication's effect had caught the defendant unaware, then his incapacity would be considered involuntary, and the medication could ground an affirmative defense. But if the defendant had been aware of the risk of intoxication and so been responsible for becoming incapacitated, then the medication's effect would provide no defense.

We might worry about the reasonable-expectation element of the tracing scheme. Tracing's extension of responsibility is usually constrained by foreseeability. We see this in Fischer and Ravizza's tracing scheme, for instance. As Vargas (2005b) objects, this constraint robs tracing of much of its explanatory promise, since many of the cases in which tracing might seem to help are cases in which the later wrongdoing was not foreseeable at the time the agent constrained her own agency. For instance, Vargas describes a manager who, as a teenager, purposefully inculcates a cool but jerky persona, and later, acting on the jerky persona, unreflectively mistreats a number of employees. Vargas denies that we can appeal to tracing to hold the manager responsible because he denies that the mistreatment was foreseeable when the manager was a teenager.

---

[106] *State v. Altimus*, 306 Minn. 462, 238 N.W.2d 851 (Minn. 1976).

What about the more standard tracing cases? Is vehicular homicide a foreseeable risk of social drinking? Fischer and Tognazzini write that "Drunk-driving cases are unproblematic precisely because everyone knows (or at least *should* know) that too much alcohol will impair the ability to drive a car" (2009, p. 532). But it being foreseeable that the agent might become too intoxicated to drive safely is not the same as it being foreseeable that the agent might nonetheless attempt to drive. Fischer and Tognazzini further explain that the foreseeability constraint does not require that you know what your wrongdoing will be "in all its florid particularity," and so I here grant tracing advocates the assumption that the agent's later, untoward behavior was foreseeable at the earlier time, at least in the central culpable-incapacity cases that motivate the addition of tracing.[107]

We also should be careful about the cases we are considering. Fischer and Ravizza's Max is supposed to be wholly incapacitated, rendered functionally insane. However, many of the culpably incapacitated agents we actually confront are only partially incapacitated. As Douglas Husak (2012) explains, the capacities to reason will often be impaired by intoxication but rarely destroyed. If the typical drunk

[107] This debate continues in the literature. Fischer and Tognazzini (2011b) have since offered a revised version of their 2009 paper, again contending that, for each of Vargas's cases, either the agent could have foreseen the wrongdoing in the right sort of way or the agent should not be held responsible. Kevin Timpe (2011) makes a similar argument, claiming that we can defuse Vargas's cases by getting a more precise grasp on the epistemic condition of responsibility. Shabo (2015) has recently offered a further argument in this thread, pointing to cases in which responsibility seems to outpace foreseeability. If tracing requires foreseeability as Fischer and Ravizza suggest, then the cases offered by Vargas and Shabo should give us concern about just how much explanatory power tracing can offer. However, because the Odysseus cases provide independent grounds for rejecting tracing, I leave the foreseeability worries aside.

driver is only partially incapacitated, then the typical drunk driver remains partially reasons-responsive. And because drunk drivers are often partially reasons-responsive at the time they drive drunk, they are partially responsible for their drunk driving as wrongdoing even without tracing. If we are not careful, our intuitions about the rare fully incapacitated drunk driver could be influenced by our experiences with much more common partially incapacitated drunk drivers. But I will assume due care in this regard.

Tracing skeptics like Matt King (2014), Andrew Khoury (2012), and Larry Alexander (2013) argue that the ordinary-responsibility account can account for cases such as Fischer and Ravizza's drunk driver even without tracing. As King explains, becoming intoxicated to the point of incapacity creates risk, and often that risk is unwarranted. When an agent creates an unwarranted risk, the agent is reckless if he is aware of the risk, and he is negligent if he is not aware of the risk but should be. It is a familiar feature of ordinary responsibility that we hold agents responsible for their reckless or negligent conduct and for the foreseeable consequences of that conduct, and so we can hold the culpably incapacitated agents responsible for their reckless or negligent conduct in becoming incapacitated as well as for the foreseeable consequences of their recklessness or negligence. We can hold drunk drivers responsible for acting improperly and for the foreseeable

consequences of that improper action without needing tracing, and so ordinary responsibility can avoid the explanatory-inadequacy worry.[108]

But tracing's advocates have insisted that ordinary responsibility is not sufficient. Timpe contends that "it is hard to see ... how one could account for a drunk driver's being responsible for running over a pedestrian without a tracing clause" (2011, p. 12). According to Fischer and Ravizza, tracing is a "refinement" developed to address a "problem" for their reasons-responsiveness theory of responsibility (1998, p. 49). And Fischer later writes with Tognazzini that tracing was a "component [that] *must be added* to get a plausible theory of moral responsibility" (2009, p. 532 emphasis added). Tracing's advocates see tracing as a necessary addition to the theory of responsibility. Merely being responsible for the foreseeable consequences of some prior action is not sufficient.

---

[108] King also suggests that tracing brings complications we can avoid if we reject tracing, citing McKenna (2008b), George Sher (2009), Angela Smith (2008), and Vargas (2005b). If foreseeability constrains responsibility, how plausible is it that the ultimate wrongdoing is foreseeable before the agent has become incompetent in the tracing cases? And what about tracing cases in which the agent has incapacitated himself thoughtlessly? Is it reasonable to hold the agent accountable for that oversight even if the agent was never cognizant of the possibility of precaution? But concerns like these cannot tell against tracing. At best, they serve to delimit the scope of tracing to cases in which the ultimate wrongdoing was foreseeable or perhaps even foreseen at the time the agent acted to incapacitate himself or failed to prevent his incapacitation. We might think that at least some cases of agents such as Max and Arsenault are like this. Further, that the tracing account must address these sorts of concerns does not give the ordinary-responsibility account any advantage, since that account must address the same sorts of concerns. Foreseeability and control matter for the ordinary-responsibility account's notion of responsibility for consequences. Responsibility for consequences is how King hopes to explain the responsibility of the culpably incapacitated actor, and so rejecting tracing does not sidestep these problems, which are really questions for accounts of responsibility more broadly. The skeptics' better argument is the explanatory-adequacy argument.

To defend this position, the tracing advocate needs to show both that tracing makes a difference and that we should want our theory of responsibility to include that difference. In the rest of this chapter, however, I will argue that tracing's advocates cannot achieve both of these goals. On the most plausible understanding of the sort of difference tracing might make, it is a difference we should reject.

4.2 Tracing and the Odysseus cases

4.2.1 The difference tracing can make

The tracing advocate needs to identify a substantive difference between the tracing account and the ordinary-responsibility account. Start by considering the formal differences. We can identify (at least) two sorts of responsibility that are formally distinguished by their objects--action responsibility and consequence responsibility--though this does not yet require that there be any substantive difference tracking this formal difference. The ordinary-responsibility account holds the culpably incapacitated agent responsible for the original incapacitating act A_1 as a bit of action[109] and for the culpably incapacitated act A_2 and any further harms (H) as consequences of the original incapacitating act:

Ordinary Responsibility: $A\{A_1\} + C_{A_1}\{A_2, H\}$

The tracing account adds to ordinary responsibility that the culpably incapacitated agent is also responsible for the incapacitated action by virtue of tracing:

Tracing Responsibility: $A\{A_1\} + C_{A_1}\{A_2, H\} + T\{A_2\}$

[109] I set aside any distinction between actions and omissions here, recognizing that responsibility for omissions presents a rich set of questions. The important distinction for considering tracing is that between actions and consequences.

But we should remember that the tracing account's extension of tracing responsibility is supposed to supplement ordinary responsibility's account of responsibility for action. That means that tracing allows us to hold the agent responsible for both the original and the later actions *qua* actions, both with their concomitant consequences:

Tracing Responsibility: $A\{A_1\} + C_{A_2}\{A_2, H\} + A\{A_2\} + C_{A_2}\{H\}$

We can see that there are two formal differences between ordinary responsibility and tracing responsibility: (a) tracing duplicates some of the ordinary responsibility's objects of responsibility, since A_2 and H each appear twice in tracing's accounting of responsibility, and (b) tracing, but not ordinary responsibility, allows us to hold the agent responsible for A_2, the culpably incapacitated action, as an action instead of only as a consequence.

Does the duplication matter?[110] It is hard to decide this without first determining whether we are right to want tracing. Suppose that the duplication leads to increased blame. If tracing's duplication leads to excessive blame, then that tells against tracing, and if tracing's imposition of culpability appears to be an instance of double jeopardy, so much the worse for tracing. But if it is true that tracing is an appropriate addition to the reasons-responsiveness scheme, then

[110] David Brink and an anonymous referee have both pressed me on this duplication. Here we see the problem that the tracing advocate faces throughout. The tracing advocate needs to find a substantive difference between ordinary responsibility and tracing that is worth having. The duplication presents an apparent substantive difference. If the tracing advocate finds some way to avoid the duplication, then the tracing advocate faces renewed pressure to find some alternative way to distinguish the tracing account from the ordinary-responsibility account.

tracing's duplication leads to the right degree of blame in the culpable-incapacity cases, and it is ordinary responsibility that has got things wrong, letting the culpably incapacitated agents off too leniently. And even this line of thinking takes as a given that the elements appearing twice heightens the degree of blame warranted, but that is an open, substantive question. So the mere fact of formal duplication does not tell against tracing.

My argument focuses instead on the second formal difference. Unlike ordinary responsibility, tracing allows us to hold the agent responsible for the culpably incapacitated action both as a consequence of the incapacitating action and as an action in its own right. But this formal difference matters only if there is a concomitant substantive difference between action responsibility and consequence responsibility. It is nearly axiomatic that being responsible for doing wrong can make an otherwise blameless person blameworthy. It can be appropriate to resent someone who has done wrong on the basis of that wrongdoing, even if they have otherwise acted appropriately, and it can be appropriate to punish someone who has done wrong on the basis of that wrongdoing, even if they have otherwise acted appropriately.[111] This exposure to blame is what it means to be accountable for a wrongful action.

---

[111] Actually, in both cases, it would be more accurate to say that we have something like a prima facie reason to resent or to punish, not an all-things-considered reason. For example, there might be reasons against resenting or punishing the blameworthy agent--perhaps the costs of resentment and punishment or the harms that might befall third parties--which make it the case that, while there is some reason to resent or to punish, all things considered it would be best not to. I set that difference aside here.

What about consequence responsibility? It seems uncontroversial that being responsible for a bad consequence matters. It can obligate an agent to make repair, and it can make it appropriate for an agent to feel a special sort of regret. Being responsible for a bad consequence can also matter for an agent's blameworthiness, though this is more controversial.[112] It can increase the degree of an otherwise-blameworthy agent's blameworthiness, and it can change the scope of an agent's blameworthiness. However, being responsible for a bad consequence cannot render an otherwise blameless person blameworthy. It would be inappropriate to resent someone who has done nothing wrong, even if they have caused harm. That is, responsibility for consequences can affect the degree or scope of blameworthiness, but it cannot by itself affect the fact of blameworthiness. Only responsibility for a wrongdoing as an action can affect the fact of blameworthiness.

Consider a surgeon who performs a risky but appropriate surgery. All surgeries carry the risk that something will go wrong, even if the surgeon takes all appropriate precautions and makes no mistakes. Sometimes things just do not work out. Imagine a surgeon who performs a consented-to, warranted surgery competently, and yet the surgery results in disaster for the patient, even the

---

[112] Responsibility for bad consequences can matter for blameworthiness even if we reject resultant moral luck. As Michael Zimmerman (2002) explains, rejecting resultant moral luck does not mean that consequences become wholly irrelevant. Even if we set aside any possible import for blame, being responsible for consequences can have other normative import. Moreover, rejecting resultant moral luck is controversial, and many reasons-responsiveness theorists accept resultant moral luck. Fischer and Ravizza, for instance, have an extended treatment of the conditions of responsibility for consequences. And it is common to see the criminal law as accepting resultant moral luck, punishing completed crimes more harshly than merely attempted crimes.

patient's death. The death was a foreseeable result of the surgeon's behavior in performing the surgery, and the surgeon was responsible for her behavior in performing the surgery. So, in some sense, the surgeon might be responsible for the patient's death. However, the surgeon is not responsible in the sense of moral responsibility I am centrally concerned with; intuitively, the surgeon is not blameworthy for the patient's death. It would be appropriate for the surgeon to feel a special sort of regret for being involved with the patient's death, and it might be appropriate for the surgeon to make some effort at repair or amends, perhaps toward the patient's family. But it would be inappropriate to blame, resent, or punish the surgeon. By contrast, consider a surgeon who performs a risky and inappropriate surgery. Luckily, the surgery is a success. Nonetheless, it seems appropriate for us to blame the surgeon. It is wrong to perform an inappropriate surgery. Being responsible for the wrongdoing is sufficient to expose the surgeon to blame, even if there are no further harms.[113]

To reprise, one formal difference between tracing responsibility and ordinary responsibility is that tracing responsibility allows us to hold the culpably incapacitated agent responsible for the culpably incapacitated wrongdoing both as a consequence and as an action, whereas ordinary responsibility only allows us to hold the culpably incapacitated agent responsible for the culpably incapacitated

[113] David Brink and Sam Rickless have pressed me to think more about the role of consent in these surgery cases. Consent seems important to the permissibility of performing surgery, but it does not seem to me to do all of the normative work. I expect the surgeon to exercise her own judgment regarding the propriety of the surgery, independent of the patient's evaluation. However, this is a complicated matter, worthy of further consideration.

wrongdoing as a consequence. But to meaningfully distinguish the two accounts, we need to identify a substantive difference tracking that formal difference. Since action responsibility can ground blameworthiness for otherwise innocent agents, tracing (but not ordinary responsibility) makes foreseeable incapacitated behavior sufficient to hold an otherwise blameless agent blameworthy. This creates the possibility of an extensional difference between ordinary responsibility and tracing responsibility.

4.2.2 Tracing gets things wrong

Recall Odysseus' encounter with the Sirens. Odysseus and his men were to sail past the Sirens on their return to Ithaca. Circe had warned Odysseus that anyone hearing the Sirens' song would be maddened by a desire to stay, never to return home. Odysseus had his men stuff their ears with wax. But Odysseus, wanting to hear the Sirens' song, had his men bind him to the ship's mast instead. When Odysseus and his men approached the Sirens, Odysseus heard their song, and he was filled with desire to stay with the Sirens. But he was bound to the mast, incapacitated, and his men would not unbind him, so he could not act upon his desire. Odysseus and his men passed safely.

Or consider a case Derek Parfit (1984) developed from Thomas Schelling's *The Strategy of Conflict* (1980). In that case, an armed robber threatens to kill an agent's children unless the agent unlocks a gold-filled safe. The agent knows that it would be irrational to provide the gold (since then the armed robber would kill the agent and her children to stop them from reporting the crime), and she also knows

that it would be irrational to ignore the threat (since that would risk the robber killing one of the children to spur the agent to action). The best choice is to take a drug, "conveniently at hand," which would render the agent irrational. The agent's irrationality would leave the armed robber's threats ineffective, since the irrational agent would no longer be moved by concern for her children. The armed robber would hopefully recognize that and decide that his best remaining option would be to escape (presumably without harming the agent or her children, perhaps to minimize his criminal exposure). As Parfit acknowledges, there is a risk that the irrational agent would harm herself or her children during the period of her irrationality. But Parfit contends that it is still appropriate for the agent to cause herself to become irrational, since that risk is outweighed by the need to defuse the armed robber's threats. As Parfit explains, "On any plausible theory about rationality, it would be rational for me, in this case, to cause myself to become for a period irrational" (1984, p. 13).

In Odysseus' case and in Parfit's rational-irrationality case, the agents use their compromised agency as a tool. Both agents solve some problem--how to experience the beauty of the Sirens' song without becoming its victim, and how to defuse the invader's threat--by giving up control. Although giving up control was risky, since things could have worked out poorly, the risk was justified.[114] And

[114] Or, perhaps more accurately, the cases are presented to us as cases in which we are supposed to take the risk to be justified. Odysseus is supposed to be a hero, and his cleverness is supposed to be his virtue. Some modern readers might find themselves less impressed with his willingness to risk the lives of those loyal to him. And some readers might likewise be unconvinced of Parfit's parent's assessment of

because both agents purposefully brought about their own incapacity, it was foreseeable to both agents that their behavior would lead to their incapacity.[115] Both Odysseus and Parfit's parent acted: a) competently, b) in a way that foreseeably led to the agent's own risky incapacitation, and yet c) morally appropriately. Call such cases *Odysseus cases*. Odysseus agents act in a way that foreseeably (and sometimes purposefully) leads to their own incapacitation, and they do so while they are competent. Accordingly, they are responsible for incapacitating themselves. Unlike the culpable-incapacity cases, however, the Odysseus agents are not blameworthy for incapacitating themselves.

Homer's and Parfit's cases are fantastic and fictional. But there are also ordinary Odysseus cases. Going to sleep presents an Odysseus case. Being asleep is risky. The sleeping agent is unaware of his surroundings, unaware of risks that might present themselves, and unable to react. But, at least in normal circumstances, those risks are slight and outweighed by the benefits of sleep. Similarly, becoming medically incapacitated is risky. Being sedated or anesthetized entails giving up control, and that presents some risk. However, anesthetic and sedation are important and valuable elements of modern medicine, and the risks they present are usually outweighed by the benefits they offer. These agents who tie themselves

---

the relative risks involved. Even if skeptical readers doubt these particular cases, they should be able to discern the pattern involved and imagine their own cases, perhaps even more fanciful, and I will shortly present more quotidian Odysseus cases.

[115] For both of these agents, risky incapacity was a tool used to achieve some goal. But this is not the key fact. We could imagine an Odysseus agent for whom the incapacity is a foreseeable side effect.

to masts, take irrationality pills, go to sleep, take sedatives, or the like willingly incapacitate themselves, but they do not do so culpably. So Odysseus cases are a feature of our ordinary lives, not merely a philosopher's construction.

The test case we need to distinguish tracing from ordinary responsibility is a special sort of Odysseus case. In addition to the incapacitation's being non-culpable, two further conditions must be met. First, unlike Odysseus himself, whose incapacitation was external, the test agent's incapacitation should be internal, arising because the agent's normative capacities are compromised. It is easy enough to imagine some medications working this way, such as Parfit's convenient pill or the physician's sedative. Second, again unlike in Odysseus' case, things have to work out poorly. In particular, there has to be some second bit of behavior, occurring during the incapacity, that is wrongful behavior.[116]

---

[116] The tracing account extends action responsibility to the later behavior. Extending action responsibility (and not merely consequence responsibility) matters for blaming only when the later behavior is wrongdoing. So, to contrast the tracing account and the ordinary-responsibility account, it is important that the second bit of behavior be wrongful behavior. In this chapter, I am agnostic as to the conditions of behavior's being wrongful. However, it is plausible that wrongful behavior requires the possession of certain mental states, and it might be that some of the conditions that mar responsibility also sometimes preclude wrongfulness. For instance, recall the example of the intoxicated burglar from n.105, where I explained that if the intoxication made it impossible for the agent to form the requisite intention, the agent did not commit burglary. Set aside those cases, and focus on cases in which an incapacitated agent can still act wrongfully.

This limitation marks a significant difference between culpable incapacity and culpable ignorance. The ultimate behavior in the culpable-incapacity cases is wrongdoing, and we are asking whether to hold the agent responsible. On a common understanding, the ultimate behavior in the culpable-ignorance cases is not ordinary wrongdoing, precisely because the agent is ignorant of some fact that bears on the behavior's inappropriateness, and we are asking whether to treat the behavior as wrongdoing nonetheless. For a clear treatment of culpable ignorance

We can imagine this sort of case by modifying the case of Parfit's self-incapacitated parent. Imagine that the robber behaves as expected, reacting to the parent's irrationality by making his escape. It takes some time, however, for the drug to wear off. In the meantime, the parent irrationally but purposefully--and therefore wrongfully--harms her children. Is she blameworthy for that wrongdoing?

Or consider a more ordinary case. Imagine a surgery patient, recovering from a desperately needed surgery, slowly emerging from the grip of a powerful anesthetic. Awake but still quite drugged, the patient mistreats the attending nurses, making repeated rude, impatient, and insulting demands. Because the anesthesia was a necessary element of a necessary surgery, the patient was properly incapacitated, even knowing that there was a good chance the patient would act impulsively while recovering from the anesthetic. In fact, the hospital requires its patients to remain under observation for a substantial period after surgery exactly because of the anesthetic's effects on appropriate judgment. Many times patients remain asleep throughout that period, but in this case the patient awoke and acted wrongly. Is the patient responsible for that wrongdoing?

Ordinary responsibility would not render these agents blameworthy. Because their incapacitating actions were justified, there is no blame to be had there. What about the incapacitated wrongdoings? These unlucky Odysseus agents are incapacitated when they act wrongfully. Because reasons-responsiveness is a necessary condition of responsibility for wrongdoing under the ordinary-

---

invoking a distinction parallel to that between the ordinary-responsibility account and the tracing account, see Holly Smith's "Culpable Ignorance" (1983).

responsibility scheme, the agents are not responsible for their wrongdoings as a bit of action. The wrongful behaviors were foreseeable in light of the agents' earlier actions in becoming incapacitated, and so they might be held responsible for the wrongdoings as consequences.[117] Hence, it might be that they should feel regret, make amends, or the like. However, as consequences, the incapacitated, wrongful behaviors cannot render the otherwise-blameless agents blameworthy.

Contrast this with tracing responsibility. Because the incapacitated wrongdoings were the foreseeable upshots of the agents' earlier behaviors, we trace responsibility for the incapacitated wrongdoings back to the agent's responsibility for their incapacitating actions. Tracing thus holds the agents responsible for their incapacitated wrongdoings as actions. Since the agents would thus be responsible for a bit of wrongdoing, the tracing account entails that the agents are blameworthy.

By considering Odysseus cases, I have identified cases in which ordinary responsibility and tracing responsibility disagree about whether an agent is blameworthy. The tracing account holds unlucky Odysseus agents blameworthy, and the ordinary-responsibility account does not. Between the two, the ordinary-responsibility account offers the more attractive verdict. Intuitively, the unlucky Odysseus agents are not blameworthy. When I imagine the modified Parfit case, for

---

[117] It might seem strange that it is true both that the later, wrongful behavior was a foreseeable consequence of the agent's earlier behavior and that the agent's earlier behavior was not wrongdoing. However, there is nothing improper about this. Lots of behavior runs risks, and so long as we think that some risks can be justified, there is room to think that a bit of behavior might not be wrong even when it results in a bad outcome. Why would anything change about this just because the bad outcome involves a risked bit of wrongdoing?

instance, I lack the intuition that the parent is blameworthy, and I lack the related intuition that she could appropriately be punished. Instead, intuitively, she seems unlucky. It is easy to imagine the parent feeling regret, and it is easy to imagine thinking poorly of her if she does not feel that regret or if she fails to make an effort to address the harms she has caused. And I could imagine feeling terrible for the parent who harmed her own child. But she does not seem blameworthy. I feel sympathy, not resentment, toward the parent.

These intuitions comport with thinking of responsibility as tracking the fair opportunity to avoid wrongdoing. Imagine holding the parent blameworthy. You can imagine her asking what she should have done differently. Should she have refrained from taking the drug, thereby exposing herself and her family to the robber's threats? Given the options available, she had no fair opportunity to avoid the risk of the wrongdoing, and so she did not have a fair opportunity to avoid the wrongdoing. Contrast this with the culpable-incapacity cases in the tracing literature, in which the agent did have a fair opportunity to avoid the wrongdoing. The drunk driver, for instance, had the fair opportunity to avoid the wrongdoing when the drunk driver had the opportunity not to become intoxicated to the point of incapacity. It is ordinarily fair to ask someone not to drink to incapacitation.

In the Odysseus cases, ordinary responsibility gets the verdicts right, and tracing responsibility gets the verdicts wrong. That and ordinary responsibility's ability to ground blame in the original culpable-incapacity cases give us sufficient reason to reject tracing and stick with ordinary responsibility.

The Odysseus cases pose a problem for the tracing account because the tracing advocate appears committed to three propositions: 1) tracing extends responsibility in cases of responsible but non-culpable incapacity, 2) tracing extends action responsibility in particular, and 3) being action responsible for a bit of wrongdoing is sufficient for blameworthiness. If those three propositions are true, then the tracing advocate is committed to holding the Odysseus agents blameworthy, and that tells against tracing. So could a defender of tracing not fend off my attack by denying one of those three propositions? Why not, for instance, limit tracing only to cases of culpable incapacity?

The problem for the tracing advocate is not just that any such limitations appear ad hoc.[118] The problem is that the tracing advocate can appeal to these responses only at the cost of making the tracing account substantively indistinguishable from the ordinary-responsibility account, and that would be to abandon a substantive account of tracing. If the tracing advocate rejects one of these three propositions, then she will run into the strongest version of the objection suggested by skeptics like Alexander, Khoury, and King, that ordinary responsibility gives us everything that tracing gives us.

---

[118] As Dana Nelkin has pointed out to me, this limitation might not appear ad hoc to everyone. Of course, the same distinction can seem ad hoc to one and principled to another. I tentatively hypothesize that this distinction is more likely to seem ad hoc to someone who thinks the manifestation of quality-of-will is the fundamental matter of responsibility and more likely to seem principled to someone who thinks that fairness is the fundamental matter, but that is a very tentative hypothesis. In any case, as I explain, even if this limitation is not ad hoc, the resulting scheme abandons any substantive distinction between tracing and non-tracing accounts, and that is the bigger problem for the tracing advocate.

Start with the possibility of limiting tracing only to cases of culpable incapacity. Tracing's advocates do not permit tracing in all cases of incapacity. For example, Fischer and Ravizza point to Roger O. Thornhill, Cary Grant's character in *North by Northwest*, who is forced to drink bourbon when his enemies want to stage a driving accident. Although Thornhill drives while intoxicated, he is not responsible for his behavior, because he is not responsible for becoming intoxicated. There is no responsibility to trace back to, and so Fischer and Ravizza limit tracing to cases of responsible incapacity.

It might seem natural to strengthen the restriction and limit tracing's extension of responsibility to cases in which the agent was not just responsible but also blameworthy for her underlying incapacity. Were tracing's application restricted in this way, tracing would not extend responsibility in the Odysseus cases because the underlying incapacitation is not blameworthy in those cases.[119]

But the tracing advocate faces a dilemma here. The tracing advocate does avoid the threat of the Odysseus cases by restricting tracing only to cases in which the underlying incapacity is culpable. But, in doing so, the tracing advocate makes tracing duplicative of ordinary responsibility. Even without tracing, the ordinary-

[119] We might think of this limitation in inheritance terms. Because the traced responsibility is rooted in the responsibility for the underlying incapacity, it might not be surprising if the traced responsibility inherited some of the features of the underlying incapacity. In the Odysseus cases, the underlying incapacitation is justified. It might seem natural in those cases to think that the traced responsibility would inherit the normative effect of the justification of the underlying incapacitation. And if that's so, then the tracing does not extend blameworthiness-grounding responsibility. I thank an anonymous referee for pointing out that this response can be thought of in inheritance terms.

responsibility account can explain why the consequences of an agent's culpable incapacity can heighten her blame, obligate her to make repair, and the like. What is left for the traced responsibility to do in such a case? If tracing is limited only to cases of culpable incapacity, then it adds nothing to the ordinary-responsibility account.

The tracing advocate will face the same objection if he attempts to avoid the threat of the Odysseus cases by denying my dialectical presumption that tracing extends action responsibility. Although the tracing advocates suggest that tracing is intended to supplement action responsibility, the conditions of extending tracing responsibility--control at some earlier point when the later wrongdoing is foreseeable--parallel the conditions required for holding an agent responsible for a consequence. And if we understand tracing as extending consequence responsibility instead of action responsibility, then the Odysseus cases pose no problem. Mere consequence responsibility cannot render an otherwise blameless agent blameworthy, and so if only consequence responsibility is extended, the Odysseus agents will not be exposed to blameworthiness. If the tracing advocate appeals to this response, however, tracing becomes substantively indistinguishable from ordinary responsibility. Even without tracing, ordinary responsibility can explain why the culpably incapacitated agent is blameworthy for the consequences of her responsible agency. If tracing does no more than extend consequence responsibility, then it adds nothing new to the ordinary-responsibility account.

Could the tracing advocate deny that being action responsible for wrongdoing grounds blameworthiness? If being responsible in this way for wrongdoing is not sufficient for blameworthiness, then there is room to hold the Odysseus agents responsible for their wrongdoing without holding them blameworthy. This strategy requires the tracing advocate to take a controversial stand on a foundational question about moral responsibility, and that should make this the least tempting distinction of the three. I suggested that it is nearly axiomatic that an agent's being blameworthy is entailed by her being responsible for a wrongdoing as an action. But not everyone accepts that responsibility for wrongdoing is sufficient for blameworthiness. For example, in a discussion with Pereboom, Fischer writes, "It is crucial here to keep in mind the distinction between moral responsibility and (say) moral blameworthiness (or praiseworthiness)" (2004, p. 157). Fischer explains that an agent's history--things like manipulation--could make it inappropriate to hold a responsible wrongdoer blameworthy. And McKenna (2012) argues that reasons-responsiveness and wrongdoing alone are not sufficient for blameworthiness; he requires the satisfaction of a quality-of-will condition in addition. Of course, it might be that Fischer's concern about manipulation and McKenna's concern about quality of will are best understood as telling against responsibility, and only thereby against blameworthiness. In any case, these particular constraints will not help the tracing advocate. There is no reason to think that all Odysseus agents are manipulated agents, and it is easy enough to imagine Odysseus agents who might satisfy a quality-of-will condition at

the time of the incapacitated wrongdoing. Nonetheless, we can see the conceptual possibility that responsibility for wrongdoing might not be sufficient for blameworthiness.

The tracing advocate here faces the same bind he faced elsewhere. In order to defend a substantive tracing account, the tracing advocate needs to identify some significant difference between tracing responsibility and ordinary responsibility. I have identified one plausible difference between tracing responsibility and ordinary responsibility, but accepting that difference tells against tracing. If the tracing advocate therefore denies that action responsibility is a distinct type of responsibility (or, at least, is distinctive in the way I have suggested), then the tracing advocate has no grounds for holding that the difference between action responsibility and consequence responsibility is more than merely formal. So the tracing advocate can deny that action responsibility is distinctive in this way only by abandoning the substantive tracing account.

This dooms tracing. The tracing advocate needs to show both that tracing makes a real difference and that we should want our theory of responsibility to accommodate that difference. However, the most promising way to distinguish the tracing account from the ordinary-responsibility account--understanding tracing as extending action responsibility--commits the tracing advocate to holding unlucky Odysseus agents responsible and therefore blameworthy. Since the unlucky Odysseus agents are intuitively not blameworthy, the tracing advocate can distinguish the tracing account from the ordinary-responsibility account only by

rendering the tracing account extensionally inadequate. The only apparent ways to defuse the threat from the Odysseus cases amount to abandoning tracing as a substantive addition. Accordingly, the Odysseus cases tell us to reject tracing as a substantive addition to responsibility.

Abandoning tracing as a substantive addition does not mean that there is no room for tracing in our thinking about moral responsibility. Even if tracing is not a substantive addition, it might yet serve as a helpful heuristic. Given the similarities between the conditions of applying tracing and the conditions of applying responsibility for consequences, we might charitably understand the arguments offered by the tracing advocates as intending to draw our attention to the role that responsibility for consequences can play in cases in which some of the consequences at issue are further actions. Indeed, philosophers working on other problems have not always treated the tracing account and the ordinary-responsibility account as distinctive accounts of responsibility. Neal Judisch (2005), for example, moves between Fischer and Ravizza's tracing account and their account of responsibility for consequences in discussing a challenge to their taking-ownership condition. And consider Vargas: "We hold someone responsible for the results of drunk driving, not because of the kind of agent they are when they get behind the wheel, but rather, because of the kind of agent they were when they started to drink" (2013, p. 273). This line of thinking allows that tracing might be an instance of ordinary responsibility for consequences, not a distinctive sort of responsibility for wrongdoing. Noticing the importance of recognizing the cases in

which an important consequence of our wrongdoing is some further wrongdoing would be an interesting result, though it would be a revision of the tracing advocates' arguments, given their insistence that the addition of tracing makes a substantive difference.

4.3 Considering three objections

I have argued that we do not need tracing to blame culpably incapacitated agents like the drunk driver, and I have argued that tracing threatens to hold non-culpable agents like Parfit's parent or the surgery patient blameworthy. But tracing has been persistently attractive, and so here I consider three worrisome objections to abandoning tracing.

First, tracing seems to make good the idea that no one should benefit from their own wrongdoing.[120] Being permitted an excuse might seem good for the wrongdoer; the excuse enables the wrongdoer to avoid blame that might otherwise be appropriate. Incapacity is the sort of condition that can ground an excuse. However, the culpably incapacitated agent brings about his own incapacity, and he does so by acting wrongfully. Permitting the culpably incapacitated agent to point to his own incapacitation as grounds for an excuse might then seem to violate the general principle against allowing wrongdoers to benefit from their wrongdoing.

---

[120] This general principle is a feature of American law familiar to many philosophers from Ronald Dworkin's discussion of the New York case *Riggs v. Palmer*, 115 N.Y. 506 (1889) in *Law's Empire* (1986). In that case, Elmer Palmer murdered his grandfather to ensure that the grandfather died before writing Palmer out of his will. As the court explained in ruling against Palmer, New York's probate law was to be interpreted in light of the general principles within the law, including the principle against allowing wrongdoers to benefit from their wrongdoing.

The culpably incapacitated agent would have garnered an ostensibly beneficial excuse, and he would have done so by acting wrongfully. At the extreme, allowing culpably incapacitated agents an excuse for their culpably incapacitated wrongdoing might even seem to give wrongdoers a strategy for insulating themselves against recrimination. As the Maine court explained in *Arsenault*:

> [T]he defense of insanity should never be extended to apply to voluntary intoxication in a murder case. It would not only open wide the door to defenses built on frauds and perjuries, but would build a broad, easy turnpike for escape. All that the crafty criminal would require for a well-planned murder, in Maine, would be a revolver in one hand to commit the deed, and a quart of intoxicating liquor in the other with which to build his excusable defense.

Accepting tracing, and thereby refusing to grant the culpably incapacitated agent an excuse, can ensure that there is no "broad, easy turnpike for escape."

We can set aside this worry. In order to know whether someone has benefited, we have to know what the relevant comparison is. In the culpable-incapacity cases, the agent's earlier, competent wrongdoing--the improper, self-incapacitating behavior--makes the agent more blameworthy than he otherwise would be. He is blameworthy for that initial behavior, and then he risks being blameworthy for further harms (including his later improper behavior) that result. That is significant blame that the agent could have avoided by not acting improperly from the outset. So the agent is not made better off by way of his wrongdoing (at least not in terms of escaping blame). Nor is the agent better off than someone who is involuntarily incapacitated. Both are similarly excused from action responsibility for their incapacitated wrongdoings, but only the culpably incapacitated agent is

blameworthy for being incapacitated. Recognizing this, we can see both why tracing might have seemed attractive in this way and why we need not actually worry about it.

Second, we might think tracing is appropriate because it seems to offer the best explanation of a comparative pattern of blaming we see in both ordinary morality and the law: We blame unlucky culpably incapacitated agents who commit some further wrongdoing more frequently and more harshly than we blame lucky culpably incapacitated agents. Consider again the drunk drivers. Drunk drivers, and especially drunk drivers who cause further harm, are exposed to significant and appropriate blame and punishment. What about agents who drink to the point of incapacity but then, luckily, neither drive while intoxicated nor cause any further harm? They are subjected to less frequent and less severe blame and punishment, both in ordinary morality and in the law.

This comparative pattern--that drunk drivers are punished more often and more severely than the merely drunk--might suggest that culpably incapacitated agents are being held responsible for their culpably incapacitated wrongdoing as a bit of action. Recall that it is responsibility for wrongdoing that is supposed to mark the difference between blameworthy and non-blameworthy agents. Since the drunk drivers and the merely drunk alike drank to the point of incapacitation, and since the merely drunk sometimes appear not to be blamed, then it might appear that drinking to the point of incapacitation is not being treated as action. Hence, the

wrongdoing that grounds the blameworthiness of the drunk drivers might seem to be their drunk driving.

That we do blame those who commit vehicular homicide more than mere drunk drivers and mere drunk drivers more than mere drunks does not mean that we should blame in these ways. Our practices are not immune to criticism and revision. For instance, Khoury and Alexander suggest that it is our competent behavior that matters for blame, not whatever follows. And so perhaps we should be blaming those who kill less than we do, though we might still expect contrition, compensation, and the like from them. And probably we should blame those who culpably incapacitate themselves more than we do. After all, drinking to the point of incapacity is ordinarily dangerous behavior. People who are that intoxicated cause a whole range of harms, and drunk driving accidents are merely one such particularly deadly result.[121] Even if, fortuitously, no further harm results, we should sanction that dangerousness. And we might accept one of these latter conclusions without committing to Khoury and Alexander's strict denial of resultant moral luck. If some such revisionary explanation is available, then the need to explain the extant comparative pattern is gone.

---

[121] As an anonymous referee noted, drinking to intoxication need not always be reckless. For instance, it is possible to imagine someone who, prior to drinking to the point of incapacitation, takes precautions to preclude later incapacitated misbehavior, such as arranging for a designated driver, handing over the car keys, or the like. Whether these precautions are sufficient to obviate the culpability for becoming intoxicated to the point of incapacity is not clear to me. However, it is true that it is possible for an agent to become intoxicated to the point of incapacity without thereby raising the sorts of risks that ordinarily render such behavior culpable. For such cautious agents, their incapacity would not be culpable, and we should not blame them, regardless of whether any harm results.

And we could, like many, accept that the results of an agent's behavior can affect the degree of appropriate blame. The results need not reflect any difference in the quality of the agent's will nor any difference in the agent's regard for others. However, the results of wrongdoing--risks imposed and harms suffered--can affect the interests of others. The culpable-incapacity cases often result in serious harms. Think of Hinckley, shot and killed by Arsenault, or think of the victims of drunk drivers. Their deaths are serious harms, and many accept that causing serious harms can render a blameworthy agent significantly more blameworthy. And, as King suggests, even culpably incapacitated agents whose further wrongdoing results in no additional harm--such as a drunk driver who fortuitously makes it home without incident--might be held accountable for the additional risk they have imposed, for the close call they created. If these harms and dangers can increase the degree of an agent's blameworthiness, then the tracing skeptic can explain why we might hold the unlucky culpably incapacitated agent far more blameworthy than we hold the lucky culpably incapacitated agent.

We can also explain why it might be appropriate to blame the drunk driver but not the agent who drinks to intoxication but luckily does not drive. We might conclude that, while both are blameworthy, it is only all-things-considered appropriate to blame the drunk driver. We see something like this in cases of de minimis blameworthiness.[122] Section 2.12(2) of the Model Penal Code, for instance,

[122] What counts as de minimis wrongdoing and whether we should withhold blame in those cases are questions I do not fully address here. For more substantive treatments, see Husak (2010) and Stanislaw Pomorski (1997).

excuses behavior "too trivial to warrant the condemnation of conviction." The de minimis defense is an element of our criminal practices and almost certainly also of our moral practices. Why might this be? Blaming and punishment are not costless. It takes effort to identify blameworthy agents, and we risk blaming and punishing the innocent. Blaming and punishment impose costs--psychological, financial, interpersonal and otherwise--on the blamer, on the punisher, and on third parties. These costs might be particularly unpalatable if the wrongdoing is fairly minor. And so we might let some wrongdoings slide, though the agents involved are blameworthy. This means that we might blame drunk drivers even though we do not blame drunks, despite both being blameworthy. The costs of blaming might be worth paying in the case of drunk drivers, while the costs might be too high in the case of merely drunks.

I do not here resolve which response the tracing skeptic should offer to address the comparative patterns in our punishing and blaming practices. However, plenty of philosophical resources can be brought to bear, from criticizing our extant practices to explaining them, none of which need tracing. Since we can comfortably address those comparative patterns without appealing to tracing, they do not pressure us to accept tracing.

Finally, rejecting tracing seems to suggest that incapacitated wrongdoings are just ordinary harmful consequences. But surely this is wrong. Both we and the agent should see the incapacitated wrongdoing as more than merely some untoward event in which the agent played some causal role. If ordinary responsibility

commands us to take this impoverished view of the relationship between the agent and the incapacitated wrongdoing, so much the worse for ordinary responsibility.[123]

But this objection to tracing skepticism arises only if we are not careful to distinguish between the many different sorts of responsibility that might be at issue.[124] As I have argued, the incapacitated agent is not responsible for the incapacitated action in a way that could ground blameworthiness. However, the agent can be responsible for the incapacitated action in other ways. We might ascribe responsibility to him in a way that permits us to make judgments about his character. For instance, we might think that the pill taken by Parfit's parent unleashed some improper impulse she otherwise would have restrained. She is not responsible for the incapacitated wrongdoing in the accountability sense, but we might make some judgment of her character because she harbored such an improper impulse at all. She is responsible for the wrongdoing in that sense, even if that is not the accountability sort of responsibility that could make her blameworthy.

Likewise, the culpably incapacitated agent might be responsible in a sense that makes it appropriate for the agent to feel regret and to make efforts at repair.

---

[123] I thank an anonymous referee for raising this objection forcefully. This objection invites tracing's advocates to say more about the boundaries of the reactions that accountability licenses, a rich area for further discussion.

[124] There is a significant literature on the many kinds of responsibility that might be at stake, from Watson's seminal "Two Faces of Responsibility" (1996) (which yields talk of attributability and accountability) to Fischer and Tognazzini's recent "The Physiognomy of Responsibility" (2011a) (where they identify a number of different attributability questions, a number of different accountability questions, and matters of responsibility that lie between the two).

Think of the lorry driver in Bernard Williams's *Moral Luck* (1982). The lorry driver faultlessly runs over a child, striking the child despite driving with due care. Though the lorry driver, by hypothesis, has done nothing wrong, we expect the lorry driver to feel a special sort of regret, and we may also think it appropriate for the driver to compensate for the harm caused. In the culpable-incapacity cases, the grounds for regret and compensation should be at least as strong. Williams's lorry driver bears only causal responsibility for striking the child. Fischer and Ravizza's drunk driver also bears causal responsibility for the harm caused; however, unlike Williams' lorry driver, Fischer and Ravizza's drunk driver is not faultless. And so just as it would be inapt for the lorry driver to think no more of the harm he caused than that it was something that happened merely in or through him, it would surely be inapt for the drunk driver to have such thoughts.

Non-blame reactions like regret deserve greater philosophical attention. However, we should distinguish them from the guilt and indignation that the reasons-responsiveness theorists and the tracing advocates take to mark accountability and blameworthiness. If we are not careful to distinguish the ways in which agents can be responsible, we might think culpably incapacitated agents are responsible simpliciter for the incapacitated wrongdoing. That could make us think we need tracing to account for culpably incapacitated agents' responsibility, and this would return us to worries about non-culpably incapacitated agents' responsibility--the Odysseus case objection. If we are careful to distinguish between the various sorts of responsibility at issue, however, we can see that tracing is not needed.

4.4 Conclusions

Tracing's advocates contend that the reasons-responsiveness account of moral responsibility needs to be augmented to account for the blameworthiness of culpably incapacitated agents. However, the ordinary-responsibility account can give us the right explanation in those cases: The culpably incapacitated agents are blameworthy for culpably incapacitating themselves. As other tracing skeptics have suggested, this defuses one motivation for adding tracing, that ordinary responsibility initially appeared to be explanatorily inadequate.

Defusing that motive, however, does not tell us that tracing gets things wrong. To show that tracing is wrong, I have offered a new argument against tracing. The addition of tracing is typically motivated by looking at cases of culpable incapacity, but I have challenged tracing by pointing to cases of non-culpable incapacity, the Odysseus cases. Tracing gets those cases wrong, and ordinary responsibility gets them right. This gives us reason to reject the addition of tracing to the ordinary-responsibility account. And I have supplemented that argument against tracing by offering explanations for tracing's continued popularity, showing how we might have been misled into thinking tracing attractive.

Rejecting tracing is no small matter. Tracing bifurcated the conditions of action responsibility, rendering an agent responsible if either the reasons-responsiveness conditions were met immediately or the tracing conditions were met historically. Rejecting tracing permits us to maintain a univocal condition of action responsibility. And, by rejecting tracing, we eliminate one historical element

of the analysis of responsibility. Without tracing, contemporaneous reasons-responsiveness is a necessary condition of responsibility. This is one step toward ascertaining just how and when an agent's history can be relevant to their responsibility for some particular bit of action.

Rejecting tracing also allows us to treat a central sort of criminal wrongdoing more honestly. It might have seemed that tracing was only an exceptional sort of responsibility. However, intoxication is involved in a tremendous proportion of violent crimes. If so many of our most serious crimes involve some degree of culpable incapacitation, it is important that we get the analysis of culpability in those cases correct. So without tracing, what should we say about intoxicated wrongdoers like Max and Arsenault? In answering that question, we will have to wrestle with difficult questions about partial responsibility, about foreseeability, and about just how risky and improper self-incapacitation is. Not all self-incapacitation, not even all intoxication, is alike. Perhaps there is a significant difference between the sort of drinking that agents such as Arsenault have engaged in--drinking far to excess, and in dangerous conditions--and the sort of social drinking that is widespread in our society. Does that difference lead to a difference in culpability? What are we to say about social drinking that unluckily leads to incapacitated wrongdoing? Given the prevalence of the behavior and the stakes of the harm involved, these pressing questions need philosophical investigation.

Chapter 4 has been adapted from Craig Agule, "Resisting Tracing's Siren Song," Journal of Ethics & Social Philosophy 10(1):1-24 (2016). The dissertation author was the sole investigator and author of this paper.

Chapter 5 Explaining the bad-history cases

Robert Alton Harris brutally murdered two teenagers in 1978. He shot them in cold blood, laughed about their killings, and then used their car to commit a bank robbery. Harris's crimes were reprehensible, and I readily respond to wrongdoings like his with disgust and resentment. Harris seems a fit candidate for blame. However, Harris's upbringing was horrific. He was neglected and abused, assaulted and institutionalized. This background deserves sympathy, and my sympathy tempers my initial resentment and disgust. Because of the close tie between appropriate resentment and responsibility, it might thus seem that Harris's responsibility for having done wrong is undermined by his bad history, even if he appears to be reasons-responsive. So our responses to bad-history cases pose explanatory-adequacy problems for the reasons-responsiveness theory of moral responsibility. In this chapter, I address that challenge.

Some argue that the bad-history cases show that we should abandon reasons-responsiveness and perhaps even compatibilism more broadly altogether. Pereboom (2001, 2014), for instance, offers a source argument to support his responsibility skepticism. We might imagine that Harris's bad history, rather than Harris, was the ultimate source of his behavior, because the bad history was the source of his character. We can imagine a similar argument about control: Harris's crimes were the upshot of his character, Harris's character was not under his control, and so his crimes were not under his control. We might then accept that the best explanation for our intuitions in these cases is that Harris is not fully

responsible because of source or control concerns. But incompatibilists like Pereboom then urge us to note that such explanations generalize. What is easily noticed in the bad-history cases is in fact true for all of us: our character and thus our behavior is in large part the product of outside forces, forces beyond our control. On this argument, there is no relevant difference between Harris and the rest of us. That should lead us to reject reasons-responsiveness and even compatibilism more broadly.

Others urge a more moderate response, allowing that we could maintain the core of the reasons-responsiveness account if we augment the theory with a history-sensitive element. For example, in response to worries about manipulated agents, Fischer and Ravizza (1998) add the taking-ownership condition of moral responsibility.[125] Recall that an agent takes ownership in the relevant sense by coming to see herself as agentially efficacious and as an appropriate target of moral assessments. Fischer and Ravizza claim that this process is the result of an ordinary moral education. An agent is then morally responsible if and only if both she is reasons-responsive and she has taken ownership for the relevant bit of her moral psychology. The taking-ownership requirement arguably allows us to identify a difference between Harris and the rest of us: unlike most of us, Harris was denied an ordinary moral education. He was abused and neglected, rather than being nurtured and taught. Accordingly, on Fischer and Ravizza's account, Harris cannot be held responsible for his behavior. But what is true of Harris is not true of most of

[125] I've already argued that we can accept much of Fischer and Ravizza's taking-responsibility account without taking on any core historical elements.

the rest of us, because most of the rest of us were given adequate moral educations. And so Fischer and Ravizza have a history-sensitive, compatibilist, reasons-responsive account of moral responsibility.

Both the skeptics and the historicists assume that the ahistorical reasons-responsiveness advocate has no good explanation for the bad-history cases. If that's so, then these critics are right: the theory must be modified or abandoned. But we should not so hastily agree with these critics, as the reasons-responsiveness advocate has much to say about these cases. The bad-history cases merit two complementary responses: bad history casts a shadow upon contemporary normative capacities, and bad history merits a sympathetic response which conflicts with blame but not with blameworthiness or responsibility. Those responses enable the reasons-responsiveness advocate to explain our intuitive responses to the bad-history cases without modifying the account of moral responsibility to include a historical element.

5.1 Robert Alton Harris's bad history

In the early summer of 1978, teenagers Michael Baker and John Mayeski planned a day of fishing, and they headed to a fast food restaurant to get lunch before leaving.[126] At the same time, Harris and his brother planned a bank robbery.

---

[126] My account of Harris's crime and his upbringing is taken from Watson (1987), the Los Angeles Times, and several of the court opinions addressing his case, including People v. Harris, 28 Cal.3d 935 (Ca. 1981) and Harris v. Pulley, 885 F.2d 1354 (9th Cir. 1989). And, of course, Harris's story is dramatic but not unique. Nelkin (2011) discusses the case of Jeremy Gross, who brutally shot and killed a store clerk during a robbery. A jury sentenced him to life in prison instead of the death penalty after hearing extensive testimony about his rotten social background.

Needing a car, the Harris brothers spotted Baker and Mayeski and kidnapped them, planning to use the teenagers' car for their robbery. The Harris brothers ordered Baker and Mayeski at gunpoint to drive to a secluded canyon in rural San Diego County. There, Harris promised the teenagers that they would go unhurt. Baker and Mayeski were instructed to walk off, wait some time, and then report the car stolen, giving a misleading description of the thieves. But then Harris shot Mayeski, first in the back, and then in the head. He chased Baker down, confronting the teenager as he cowered in a bush and telling him to "quit crying, and die like a man" before shooting him four times. Harris shot Mayeski again, point blank, with his pistol, before taking a rifle dropped by his brother and shooting Mayeski one final time. Harris later laughed at having shot Baker's arm off, he laughed at the idea that the brothers might pose as police officers and inform the teenagers' parents that their sons had been killed, and he laughed as he flicked bits of Mayeski's flesh off of his pistol. Harris coolly ate a carryout hamburger the teenagers had left behind, scoffing at his brother for failing to join him. Harris and his brother would go on to commit the bank robbery before being captured by police.[127] Assuming that Harris was responsible for his actions, then Harris was tremendously blameworthy and we have great reason to blame him.

For a moving account of the emotional difficulty of serving as a juror in such a case, see Alex Kotlowitz's "In the Face of Death," *The New York Times Magazine* (July 6, 2003).

[127] Horribly, one of the police officers who arrested Harris later that day was Detective Steven Baker, Michael Baker's father. At the time, Baker had no idea his son had been killed.

But Harris's blameworthiness is not all of the story. Thinking of a suggestion from Strawson, Watson asks what we are to make of Harris's "'being unfortunate in formative circumstances'" (1987, pp. 271–272 quoting Strawson). Harris's father, a decorated World War II veteran, suffered from shell shock, and his mother grew up in severe poverty. Both abused alcohol. Harris's father viciously mistreated his entire family, physically abusing all of them and sexually assaulting Harris's sisters. Harris was born months premature after his father, intoxicated and questioning his wife's (Harris's mother's) fidelity, attacked her, sending her into labor. She would later say that bringing Harris home from the hospital was like "taking a stranger's baby home." Harris's father never accepted Harris. He beat Harris with a bamboo cane, and he threatened to shoot Harris, loading his gun and telling Harris to run. Harris's mother came to resent Harris, perhaps because of the abuse she herself suffered and the poverty under which the family labored. When Harris's father was eventually jailed for sexual abuse, Harris's mother took Harris's siblings and left, abandoning Harris at 14. Harris's mistreatment by his family was exacerbated by his experiences at school. He suffered from a learning disability and from a speech problem, which led to teasing and self-doubt. However, there was no money for treatment. Instead, Harris spent most of his formative teenage years incarcerated in youth facilities and prisons, learning to fight, becoming meaner. While confined, Harris was raped several times, and he twice attempted suicide by slashing his own wrists.

These two aspects of Harris's story commonly lead to conflicting reactions. When I focus on Harris's crimes and the suffering he caused, I see him as blameworthy. But when I learn of Harris's childhood, when I imagine those conditions, I feel sympathy toward him. Even as then-Governor Pete Wilson denied Harris clemency, he noted his great compassion for "Robert Harris the child." And my sympathy tempers my resentment. This is how McKenna sees Harris's case: "The modification of our antipathy can be understood as a psychologically unavoidable effect of learning of Harris's past" (1998, p. 138). This is the phenomenon to be explained.

5.2 Harris's history-compromised reasons-responsiveness

The reasons-responsiveness account excuses agents whose reasons-responsiveness capacities are compromised, and the various component capacities of reasons-responsiveness mark various excuses the account can comfortably explain. Reasons-responsiveness and its components are features of an agent's normative psychology, and of course, an agent's normative psychology has a causal history. Accordingly, an agent's history, be it good or bad, can ground an excuse when the history "casts a shadow" on the present by resulting in compromised reasons-responsiveness. This much should be uncontroversial, but attending to the force of the threat bad history poses to reasons-responsiveness makes clear just how powerful an explanation shadow-casting is for the bad-history cases.

5.2.1 The shadow cast by bad history

Although the study of the long-term impact of juvenile trauma is relatively new, and although I should be cautious translating from the lessons of the sociologists and criminologists to the concerns of the philosopher, the empirical literature offers support for the intuition that childhood trauma might compromise reasons-responsiveness.[128] Empirical work confirms that children exposed to trauma later display a range of disorders, anxieties, and phobias. As psychiatrist Bruce Perry writes, "Traumatic experiences in childhood increase the risk of developing a variety of neuropsychiatric symptoms in adolescence and adulthood"

[128] See, e.g., Bruce Perry et al. (1995), Robert Pynoos et al. (1999), Christine Heim and Charles B. Nemeroff (2001), Virginia Ann De Sanctis et al. (2012), and James A. Reavis et al. (2013).

(1995, p. 273) And social scientists have adduced evidence that early trauma can, at least in part, explain later aggressive behavior and criminality.

There are a number of plausible accounts of how early trauma might yield later problems. Christine Heim and Charles B. Nemeroff describe the effect of childhood trauma on the nervous system's use of corticotropin-releasing factor (CRF). As they explain, CRF "is generally acknowledged to be the major coordinator of the behavioral, autonomic, immune, and endocrine components of the mammalian stress response" (2001, p. 1024). Corticotropin is often produced as a response to stress, and one of its primary functions is to spark production and release of cortisol. Problems with CRF distribution are associated with major depression and with anxiety disorders, such as post-traumatic stress disorder. But trauma during development seems to mar the construction of healthy CRF mechanisms. For instance, even a day of separation from their mothers saw a tremendous reduction in pituitary CRF-receptors in young rats, likely a response to CRF-overstimulation. Likewise, overexposure to stress may lead to reduced hippocampal volume and to alterations in the prefrontal cortex and the amygdala. These regions of the brain are believed to be particularly important for decision-making, reward-assessment, and dealing with anxiety. Thus, the impact of stress and trauma on the developing brain is both wide and deep, spanning much of the brain and causing serious changes.

Or consider the evidence that Perry (1995) offers of the manner in which early trauma affects the development of neural pathways. Neurons adapt to the

signals they receive; the more frequently a neural activation occurs, the more permanent a mark is made. Relatedly, a regular pattern of activation can result in an increase in sensitivity. And in the developing brain, repeated exposure can change the characteristics of the later adult brain. The lack of appropriate experiences during immaturity or significant mistreatment can lead to "major abnormalities or deficits in neurodevelopment--some of which may not be reversible" (Perry et al., 1995, p. 276). Perry points to significant examples of developing treatment-resistant attachment problems, including, "often, the remorseless, violent child" (1995, p. 277).

All of this evidence can explain how the very young respond to trauma. As Perry explains, each time a child is exposed to trauma, the child's developing response mechanisms are activated. With repeated activation, the mechanisms can become hyper- or hypoactive; this is the upshot of the neural learning just described. The child can become more sensitive to trauma, leading to even more activation. This process is exacerbated if the child's ordinary tools for dealing with trauma are themselves dysfunctional. For instance, unable to defend themselves, infants and young children appeal to their protectors (usually their parents) in times of trauma. A successful appeal both abates the trauma and binds the child to the parent. If the appeal is unsuccessful, however, the originating trauma is compounded by the alienation from the parent and the sense that the child can do nothing to abate the trauma.

This evidence of brain mechanism and traumatic response provides support for the intuitive claim that developmental stress mars psychological health and function. Childhood stress is associated with post-traumatic stress, attention deficit/hyperactivity disorder, depression, and mood and anxiety disorders. As Perry explains, repeated trauma during childhood can lead to a long list of ills: "motor hyperactivity, anxiety, behavioral impulsivity, sleep problems, tachycardia, hypertension, and a variety of neuroendocrine abnormalities" (1995, p. 278). The brain regions affected by stress are important for regulating arousal and affect. Because of this, a repeatedly traumatized child may struggle to differentiate and understand his emotions, and he may fail to develop normal control of aggression and assertiveness. Trauma-response emotions like courage and fear may be suppressed or heightened. These emotional effects may interact with cognitive disruptions, as traumatic experiences can mar the capacities to properly assess risk and danger. Then there are psychological effects which develop as a defensive response to trauma. For example, a child who regularly witnesses others subjected to trauma may develop an empathy deficit as a protective defense mechanism. There's good reason to believe that these dysfunctional and defensive effects may be additive, such that a child who is repeatedly exposed to trauma will experience cumulative harms.

It isn't clear that the philosophers and the scientists are deploying the same terminology and concepts, hampering the easy translation of the empirical research into philosophical implications. However, it does seem that early trauma can

undermine the reasons-responsiveness capacities. Begin with the cognitive capacity. Even in this brief survey, there are a number of ways that a bad history may cause problems for the agent's later capacity to recognize the reasons there are and understand how they fit together. Recognizing reasons requires attending to them, and so attention deficits compromise the cognitive capacity. And risk-assessment problems are also cognitive problems. In almost all cases, the propriety of some action depends upon the risks attending to the choice. How likely are benefits to accrue? How likely are harms to result? If the agent has difficulty assessing risks, then the agent will struggle to correctly evaluate the choices she faces. For example, an agent who has become desensitized to risk may be unable to correctly assess the risks to others her choices would create, and this may lead her to impose undue risks on others.

Childhood trauma also undermines the development of the volitional capacity. Many of the emotional effects associated with a traumatic upbringing bear directly on the capacities needed to properly convert practical judgments into actions. Hyperactivity, impulsivity, and inadequate control of aggression and assertiveness all compromise that functioning. Of course, the possibility that an agent might act occasionally on impulse does not render the agent irresponsible. But in the bad-history cases, where impulse control is pathological, we should more readily believe the agent's volitional capacity has been compromised.

5.2.2 The mixed evidence Harris presented to the courts

We can now revisit Harris's case to see how a closer look at a particular bad-history case can be expected to yield evidence of compromised reasons-responsiveness. Although Harris plainly did not escape his upbringing with a healthy normative psychology, we have only limited information about his exact mental condition. What we know about his condition is limited to what was revealed in the course of the courts' complicated adjudication of Harris's crimes and punishment. At trial, Harris's counsel did not present evidence regarding Harris's mental health, instead blaming the killings on Harris's brother. Harris's counsel thought that a mental-health defense would be seen as inconsistent with this on-the-elements defense. However, Harris's mental condition was at issue during the sentencing phase of his trial. In order to undermine Harris's belated claim to regret the killings, the prosecutor introduced evidence that Harris suffered from antisocial personality disorder. A prosecution psychiatrist testified that, while those suffering from antisocial personality disorder were likely to regret being caught and punished, they were unlikely to experience remorse for their wrongs. The argument was that, if Harris lacked the capacity for remorse, his testimony at sentencing that he regretted the crimes should be dismissed or even held against him as dissembling. The psychiatrist also testified that people who suffer from antisocial personality disorder "are immature, emotionally unstable, they're callous, rather rigid at times, they're irresponsible, impulsive, egotistical, somewhat passively aggressive at times, they seem to have an inability to profit from past experience or

punishment." The psychiatrist explained that Harris's antisocial personality disorder likely arose because of Harris's mistreatment as a child.

Harris's mental condition at the time of the crimes was brought up again during his appellate and post-appellate litigation, especially in the litigation over the effectiveness of his trial counsel. In addition to affirming the prosecution psychiatrist's diagnosis, Harris's later doctors also found evidence of a range of other problems, including fetal alcohol syndrome, post-traumatic stress disorder, childhood head trauma, and organic brain damage. Like the prosecution's doctor at sentencing, these later doctors traced Harris's conditions to his mistreatment early in life.[129]

This further psychiatric evidence, however, proved no aid to Harris. The appellate courts that visited the issue considered the evidence procedurally suspect.[130] But even setting aside those procedural concerns, the courts were skeptical that Harris's psychiatric evidence should excuse. In one ruling, the Ninth Circuit Court of Appeals commented that Harris's conditions were distinguishable from "incurable and dangerous" conditions like paranoid schizophrenia and

---

[129] Tellingly, one doctor explained that Harris's mental health seemed to have improved on death row, crediting Harris with using "the opportunity of a highly structured and relatively low stress environment of death row to weed out much of his PTSD symptomology and related behavior."

[130] For instance, one appellate court ruled that Harris had forfeited the right to present additional psychiatric evidence because nine years had passed before he presented his evidence. The court explained that the years-long delay had prejudiced the state's ability to respond, noting that one of Harris's examining physicians had passed away during that time.

hallucinations.[131] The court worried about treating more leniently someone who rejected society's social norms (perhaps as an element of a mental condition) than someone who accepted them generally but flaunted them on a particular occasion, and the court worried that there might be some diagnosable condition involved in every crime as vicious as Harris's. Most relevantly, the court pointed out that Harris's conditions did not leave him unable to understand the consequences of his actions and that Harris's condition did not leave him unable to refrain from performing the particular crimes.

But we should not defer to the decisions of the courts, which face evidentiary and procedural hurdles not binding on philosophical investigations, and we should distinguish Harris the defendant from Harris the intuition pump. While the psychological capacities of Harris the defendant are a matter of fact we might investigate by turning to the extensive record generated during his trial and subsequent litigation, the philosophical problem arises because of our conflicting responses to Harris's case as a limited vignette. Accordingly, we may fairly confine ourselves to that vignette. What can the reasons-responsiveness advocate say about Harris the intuition pump?

5.2.3 Harris's compromised risk sensitivity

Harris's crimes were multiple, complex, and connected, and most of the wrongdoing was planned in advance. He and his brother spent a great deal of time

---

[131] The court's concern with the potential for cure is misplaced. That Harris could have cured his condition does not mean that the failure to cure the condition renders the condition normatively irrelevant, especially as we should reject the tracing doctrine. The concern with dangerousness, however, is more important.

plotting their bank robbery, considering the risks, and preparing for the crime. But at the core of his culpability is the murder of the two teens. Unlike the bank robbery and the theft of the car, the murders seem to have been spontaneous and unplanned. Harris's brother testified (albeit self-servingly) that he was surprised by the killings. I suspect that Harris killed the two teenagers as a last-second decision when he feared that they would turn the brothers in to the police. While imprisoned, Harris told another inmate: "I couldn't have no punks running around that could identify me, so I wasted them." Thus we can understand the murders as an immediate response to a perceived threat.

But Harris's rotten social background almost surely marred his ability to assess and respond to threats. The healthy exposure to and defusing of apparent risks during childhood is an important element of developing a healthy cognitive faculty for assessing risk during adulthood. Because Harris was repeatedly exposed to traumatic risks, and because he was almost never appropriately protected from those risks, we can be confident that Harris's cognition of risks was compromised. In particular, we should expect that Harris was hypersensitive to risks. As Perry writes, for those subjected to serious developmental trauma, "full-blown response patterns … can be elicited by apparently minor stressors" (1995, p. 275). Children who have undergone serious trauma "are hyperreactive and overly sensitive … very easily moved from being mildly anxious to feeling threatened to being terrorized" (1995, p. 278). We can connect these effects to the two reasons-responsiveness capacities. We should expect that an agent in the throes of oversensitivity to risk will find it

difficult to notice countervailing matters, and that cognitive effect will likely be paired with volitional impairment. The repeated exposure to trauma likely undermined Harris's restraint and his control of his aggression, and it likely left him more exposed to impulsive behaviors.

These issues should not affect Harris's responsibility for the bank robbery and the car theft. While Harris may not have been in possession of an unblemished faculty of risk perception, much of that compromise was likely compensated for by Harris's advanced planning, which left him able to coolly consider the consequences of his behaviors and able to plan in light of perceived risks. The car theft and the bank robbery were not wrongs of impulse.

But we can understand Harris's killing of the teenagers differently. Of course, the risk that the teenagers would turn him in to the police would not itself ground any justification or excuse, partial or otherwise. The risk of being reported is not the sort of factor that could support a killing (at least in ordinary cases). Nor would the mere misjudgment of that risk excuse. If a higher risk of getting reported would not justify the killing, then mistakenly judging there to be a higher risk of getting reported would not excuse either. However, as the murders seem to have been rash and unplanned, Harris's compromised ability to assess risk and his compromised impulse control surely played a role in the killings. It is plausible both that Harris overestimated the marginal effect on his risk of being caught and that Harris overweighed the pragmatic downside of being identified. These misperceptions, combined with his likely compromised volitional capacity, likely made it

correspondingly more difficult to control the impulse to kill the teenagers. If that's so, we might reasonably conclude that Harris was not fully reasons-responsive with respect to the murders.[132]

That Harris's impulsivity might have limited his reasons-responsiveness with respect to the murders is an important conclusion. Though Harris committed a number of crimes during his spree--the car theft and the bank robbery among them--the two murders are clearly the worst of the lot. Our reactions to Harris are largely our reactions to the two murders. So even if he was fully responsible for every other element of his spree, we should expect his compromised responsibility for the murders to significantly affect his overall level of culpability. Because of this, an intuition that Harris is not fully culpable can be explained as tracking his expected compromised responsibility for the two murders. His bad history indirectly explains why he is not fully responsible for the murders. So here we have one explanation for our experienced response to Harris's case that gives his history an indirect role.

5.2.4 Harris's compromised empathy

My second argument is that Harris's traumatic upbringing compromised his capacity for empathy and that this compromised empathy limited Harris's capacity to discern, make sense of, and respond to the interests of the teenagers and the

---

[132] Of course, that Harris's impulse control was compromised does not mean that it was so compromised that he might be entitled to an excuse. Perfect impulse control is not required for responsibility, only substantial impulse control, and it is not clear that Harris lacked substantial impulse control. This is a place where further investigation into Harris's particulars might help inform the debate. As I explain shortly, however, even the mere suspicion might help explain our discomfort with Harris's case.

reasons not to kidnap and then kill them. As I explained previously, empathy can play a number of roles in the reasons-responsiveness scheme. In development, empathy can help habituate us to respond to moral reasons, and at the time of our behavior, it can help to draw our attention to certain moral reasons and it can help motivate appropriate behavior. Moreover, if a thick understanding of the relevant moral reasons is required, empathy might even be necessary for cognition of at least certain reasons. These roles for empathy are not wholly uncontroversial. The constitutive role for empathy (the last role described) in particular requires a controversial account of the sort of knowledge required for moral responsibility, and the other roles for empathy (where empathy serves as an aid to the responsibility capacities) may be helpful but not necessary for moral responsibility.

While the philosophical role for empathy might be controversial, it is almost certain that Harris's capacity to empathize was drastically compromised. The reports from Harris's family, from those around him in prison after his incarceration, and from the physicians who testified throughout his legal proceedings make it reasonable to conclude he had severe empathy problems. Harris's mother explained that "The only way he could vent his feelings was to break or kill something. ... He had no feeling for life, no sense of remorse." And one of his jailmates commented of Harris that, "He doesn't care about life, he doesn't care about others, he doesn't care about himself." Harris is like many bad-history agents in that his bad history apparently compromised his capacity for empathy.

Assuming that Harris was burdened with an empathetic deficit, it would have been marginally more difficult for him to see that the killings would have been wrong, and it would have been marginally more difficult for him to resist the pragmatic urge to kill the teenagers. We should expect that his cognitive and volitional capacities were less sensitive, and we should expect that they functioned without the ordinary aid of empathy. And Harris might have been denied the possibility of seeing the nature of the wrongness of the killings altogether. The depth of his empathy deficit combined with the richness of the roles plausibly played by empathy should make us reticent to conclude that Harris possessed the requisite reasons-responsiveness capacities. And that reticence can play a role in explaining the felt discomfort with judging Harris fully blameworthy.

These are thus two ways we might suspect Harris's rotten background might ground excuse. Neither explanation, however, requires that Harris's history be given a direct role in the determination of his moral responsibility. Instead, in both cases, Harris's history has an indirect impact on his responsibility. Moreover, it is plausible that neither of these potential excuses will be availing. Harris did have substantial control over his behavior, and the role of empathy in moral responsibility is controversial. However, explanations like these might accord with the thought that Harris's background was so unlike ours that surely it is likely that he did not escape his childhood relevantly unscathed. That is, these potential explanations can highlight a background suspicion we should have about Harris. These explanations and that background suspicion can explain why our intuitive response to Harris's

case is sensitive to Harris's history without giving history a direct role in the determination of responsibility.

5.3 Being sympathetic to bad-history agents

In many bad-history cases, we should expect that the agent's bad-history has "cast a shadow," leaving the agent without the full reasons-responsiveness capacities required for moral responsibility. But we do not need to look to compromised responsibility to explain why we should refrain from fully blaming the bad-history agents. It is because we should also feel sympathy for the bad-history agents that we should not fully blame them. The bad-history agents are both blame- and sympathy-worthy, and so we have pro tanto reason to blame and pro tanto reason to sympathize. However, in many cases, we cannot fully make good on both of those warranted responses--we cannot both fully blame and fully sympathize. Thus the grounds for sympathy (and not merely what those grounds might tell us about the agent's reasons-responsiveness) tell against fully blaming the bad-history agents.

This insight is of tremendous importance. Many of the most serious cases of wrongdoing involve bad history, and so we need to make sense of the relationship between blame and sympathy to get our responses right in those cases. And this insight is also important in addressing historicism about responsibility, given the centrality of the bad-history cases for many historicists' arguments. The historicists think that a history-sensitive component of responsibility explains our intuitions about those cases. My argument here provides a compelling, alternative explanation,

defusing the usual arguments for historicism, and strengthening the case for an ahistoricist, reasons-responsiveness account of responsibility.[133]

5.3.1 The distinctive conflicts in overlap cases

To tell that story, I begin by looking at a class of cases I call "overlap cases." In overlap cases, the grounds which make two (or more) distinct responses fitting are co-present. Consider the charming colleague case offered by Wallace (1994, pp. 76–77).[134] This colleague is especially charming; perhaps this colleague is attentive and witty, or perhaps he is an engaging and sympathetic listener, a confident but not dominating interlocutor. However, the colleague has also cheated and lied to you. Blame is a fitting response to the colleague's wrongdoing, but being charmed is a fitting responsive to the colleague's charm. In some overlap cases, like Wallace's charming colleague case, the two (or more) fitting responses are both directed at the same agent. But other overlap cases are more complex than this. In Harris's case, for instance, blame is fitting toward Harris but blame is also fitting toward those who neglected and mistreated Harris. What is important is there are simultaneously two or more responses it would be fitting for the evaluating agent to experience. Overlap cases of both sorts are neither rare nor unfamiliar. They are the ordinary cases. Because the world is complicated, overlap cases abound.

---

[133] McKenna (2012) gestures at a similar argument.

[134] Wallace uses this case to distinguish between judging someone blameworthy and blaming them. We judge the charming colleague blameworthy without blaming him. We judge the colleague blameworthy because we judge that the colleague was responsible and committed a wrongdoing. But because we don't experience the requisite affective response, our response to him does not include the blaming reactive attitudes.

Overlap cases yield a distinctive conflict. Consider the charming colleague. There, as Wallace explains, the colleague's charming nature makes it difficult to work up full resentment. It is difficult, and perhaps even impossible, both to be fully charmed by the colleague and to fully blame the colleague. This conflict is an ordinary phenomenon.

We should expect conflicts like this because the responses require resources, psychological resources in particular, and we are finite agents, with finite resources. For instance, our reactive attitudes both require and command attention. It is familiar that noticing, focusing, and dwelling on a wrong naturally lead to blame and resentment, and it is familiar that the stronger the attention, the stronger the blame experienced. Correspondingly, it should not be surprising that we might fail to muster up full resentment unless we sufficiently attend to the wrongdoing at issue. So our responses seem to require attention. And our responses can capture our attention. This is one important consequence of the seeing-as dimension of the reactive attitudes. When we resent someone, our resentment focuses our attention on the wrongdoing at issue, we see the wrongdoer in light of the wrongdoing, and we more readily notice matters consistent with seeing the wrongdoer's behavior as expressing ill will and more readily overlook matters inconsistent with seeing things that way. This explains why it is a familiar experience that it is difficult when angry to attend to other affairs. When we are angry, we are highly aware of the nature of the wrong at issue, and we often think and examine the wrong mentally, considering its causes, its effects, and the blameworthiness of the wrongdoer. This effect

occludes our attending to other matters, including the grounds for other responses. Because our attention is limited, our responses are limited. We can only attend to so much, and so we can only respond to so much.

Likewise, many of our responses are associated with distinctive phenomenologies. Anger has a feel, and being charmed has a feel. Perhaps those phenomenologies are constitutive of the various responses, perhaps they are causal upshots of the responses, or perhaps there is some other relationship. In any event, the phenomenologies are central to the responses. So to be angry seems to require the distinctive phenomenology of anger, and likewise for the other responses. Here too we see reason to anticipate conflict in the overlap cases. We can experience only so much at a time, and in fact we can experience only very little at a time. For instance, it would be difficult to feel both resentment and gratefulness at the same time. Because our experiences are limited, our responses are limited.

These conflicts are complicated when we recall that reactive attitudes might be experienced in either episodes or stances. It seems plain to me that the episodes of different responses might conflict with each other. It is less clear how stances might conflict, or whether we could make sense of a conflict between a stance of one reactive attitude and an episode of another. I set those complications aside, thinking from here about episodes of resentment, sympathy, being charmed, and the like.

Return to the charming colleague. When we are charmed, we focus on the charmer's flattery and wit. We see the charmer in light of his charming behavior, interpreting the charmer's other features positively and sympathetically, and

discounting features inconsistent with being charming. This framing is in tension with attending to the charmer's wrongdoing. While we might in some sense be simultaneously aware that the charming colleague is witty and a wrongdoer, we can only attend to so much at a time, and so we cannot simultaneously be fully charmed and fully resentful. And being charmed requires a certain warmth, notably distinct from the heat of anger. We cannot simultaneously experience both. In both dimensions, blame conflicts with being charmed.

These conflicts have important normative upshots. We might, as a standing assumption, take it that we have some reason to experience fitting responses.[135] In the overlap cases, this means that we have some reason to experience at least two different responses. By the instrumental principle, we also have some reason to engage in the necessary conditions of those responses. However, because of the conflicts between the responses, the necessary conditions of any one of the responses will include foregoing or at least curtailing the other responses. But of course, this means that we have reason to engage in explicitly conflicting behaviors. In order to figure out what responses we should have, then, we would need to reflect on the values of the responses and the practical limits of our behavior.

Consider two ways of developing Wallace's charming-colleague case. In the first variant, the wrongdoing committed by the colleague is slight. In that case, the lesser nature of the wrongdoing reduces the importance of blaming the colleague.

---

[135] As I noted before, this assumption is widely assumed and sometimes explicitly stated. McKenna (2012, p. 36), for instance, claims that blameworthiness yields a pro tanto reason to blame. But it is rarely defended.

The grounds for blaming being therefore weaker, dwelling on it in order to avoid the eclipsing effect of the colleague's charm might be inappropriate or at least not required. In the case where the wrongdoing is slight, therefore, there might be nothing inappropriate about failing to blame, even though the charming colleague is blameworthy. In the second variant, by contrast, imagine that the wrongdoing committed by the colleague is a significant wrongdoing to some innocent and seriously harmed third party. This might make the case for blame especially pressing. It is still of course true that fully blaming the agent requires avoiding or overcoming the urge to be charmed, though perhaps the attention-grabbing viciousness of the wrongdoing will make this easier. But the costs required to avoid being charmed seem worth paying.

Of course, we might avoid these conflicts if our psychologies were such as to express both or all responses comfortably. It might sometimes be open to an agent to change her moral psychology and to thereby lessen the degree to which her responses come into conflict. Even if this is not true immediately, it might be true in the longer run. But this too has significant costs. It is not free to change one's psychology. It might require a prolonged and difficult effort. And there's no reason to think that healthy, virtuous, duly responsive human agents won't be marked by something like our ordinary psychological conflicts. Perhaps more sophisticated agents, non-human agents, might escape these conflicts. But for us, there is good reason to think that we should have some conflicts. We need not think that the conflicts which we happen to have now are exactly those which it would be best to

have. But it is probably true that there are some conflicts which we should have, given that we are finite beings, and hopefully our actual conflicts are not too far from those. If that's so, then the right arrangement of our limited faculties is one which would in fact constrain the responses we can have at the same time.

I should be careful about the implications of the reasons to blame being outweighed. First, that the reasons to blame are outweighed can sometimes mean that it is inappropriate to blame. If those outweighing reasons are important enough, failing to heed them would be wrong, and blaming would psychologically entail failing to heed those reasons. Thus, if the outweighing reasons are important enough, blaming could be inappropriate. But that the reasons to blame are outweighed need not mean that it is inappropriate to blame; it might merely mean that it is not inappropriate not to blame. Whether it is inappropriate to blame or just not inappropriate not to blame might depend, for instance, upon the degree to which the costs of blaming would be borne by the would-be blamer. For instance, because blaming can commandeer attention, one cost of blaming is the disruption it can cause to one's other projects, and this is a cost ordinarily borne by the blamer. When the costs of blaming would largely be borne by the would-be blamer, it might be thought that whether to give effect to the normative weight of those costs is up to the would-be blamer. If that's so, blaming might be permissible, but not blaming might also be permissible.

Second, that it might be inappropriate to blame an agent does not entail that the agent blamed has grounds to object to being blamed. Not everyone has the same

right to object to a given wrong, and for some wrongs, some people might have no right or a de minimis right to object. Whether an agent has grounds to object to a particular wrong depends upon whether their interests are at stake in the wrongdoing. Sometimes the blameworthy agent has important interests in the conflicting reactions in the overlap cases, but not always. Think again of charm. We have an important interest in being perceived as charming--it is easy to imagine my life being better in important ways if others (correctly!) take me to be charming. However, there are also significant reasons to respond to charm which lie in other agents. It is pleasant to be one who is charmed. Where the reasons supporting the attitude which conflicts with blame are significantly tied to the interests of agents other than the agent who would be blamed, that agent has correspondingly weaker reason to object to being blamed.[136]

Third, the outweighing effect need not always be binary. Sometimes the outweighing costs might render it inappropriate to blame altogether, but sometimes they might merely render it inappropriate to fully blame. Whether that is so depends upon whether mixed responses are possible. Plausibly, we can experience some degree of blame and some degree of some other response. If that's so, then the outweighing costs could make some mixed response a permissible or even preferred response.

---

[136] This is perhaps easier to see in cases where the reasons against blaming a blameworthy agent are more ordinary consequences. For instance, imagine a case where the reasons to blame some agent are outweighed by the harms that blaming the agent would impose on innocent third parties. This might be a case where blaming a parent harms their child. In that case, even if it is inappropriate to blame the parent, the right to object lies in the child, not in the parent.

Finally, even if blameworthiness is outweighed, that does not entail that blameworthiness becomes insignificant. Blameworthiness can serve a number of roles. One role is to ground a pro tanto reason to blame, and my argument here suggests that when blameworthiness is outweighed that pro tanto reason does not yield an all-things-considered reason. But blameworthiness might still have other roles. Blameworthiness might, for instance, make it appropriate for us to adjust our interpersonal relationships, or blameworthiness might be seen as grounds for making certain sorts of inferences about an agent's character. Just as I mean to focus on the sort of responsibility associated with the appropriateness of blame, and not other sorts of responsibility, I mean to focus here on the role of blameworthiness in rendering blame appropriate, and not any other roles for blameworthiness. I set considering those roles aside for some other time.

My discussion of the normative features of the overlap cases takes me outside the core fittingness concerns of a theory of moral responsibility. Why not think that, in such cases, our theory of blame might rightly say only that blame is appropriate as far as it is fitting and leave it to our comprehensive moral theory to say that blame is inappropriate all-things-considered? Forcing a theory of appropriate blame to address every fact which might bear on the all-things-considered appropriateness of blame seems to preclude having a theory distinctively about blame. In that case, there would just be the larger theory of the reasons that make behaviors appropriate or inappropriate, applied to the particular

case of blame.[137] So it might seem that we should cabin our theory of appropriate blame to internal matters, matters like quality of will and reasons-responsiveness.

Although I have sympathy with this concern, it is not fatal to my argument. Suppose that we accept that our moral responsibility theorizing should be limited to the internal norms of blame. Still, a theorist concerned only with that more limited project should take heed of the possibility of outweighing reasons in the overlap cases. So long as we appeal to our intuitions about cases where blame's appropriateness is at issue in order to pick out the content of a theory of blameworthiness, we need to be aware that there could be factors rendering blame inappropriate beyond the lack of blameworthiness. We need to avoid the moralistic fallacy. And so a theorist who fails to consider those latter situations might produce a theory which errs, attributing some externally driven phenomenon to internal features for want of broader consideration.

And the overlap cases aren't just any old phenomenon which an overarching moral theory should account for. In discussing the relationship between blameworthiness and the appropriateness of blame, McKenna offers an example where blaming would result in the destruction of the planet. While that consequence surely has normative significance, it is only contingently and remotely related to blame. There's nothing particular about blame that makes it the case that

---

[137] Vargas argues that "a theory of responsibility must be silent on those considerations whose origin or normative force places them external to the norms of responsibility" (2013, p. 184). Once we go beyond matters internal to the propriety of blame, "the question becomes uninteresting for a theory of responsibility" (*ibid.*).

one instance of blame threatens the planet. By contrast, the concerns posed in the overlap cases are intimately concerned with the nature of reactive attitudes. The countervailing reasons of the overlap cases arise because of the phenomenological experiences and the framing effects of the reactive attitudes. These overlap cases arise because of distinctive features of the reactive attitudes, and surely a theory of the conditions of the appropriateness of the reactive attitudes should attend to their distinctive features.

Finally, my concern is not solely with a theory of responsibility or a theory of blameworthiness. My interests are decidedly practical. I am concerned about getting blame right. A theory of the norms internal to blaming is important to that practical project. But I also want to understand how blame ordinarily operates and what import blame has for blamers like us and wrongdoings like ours. So while a theory of the internal norms of blame might well omit any reference to countervailing reasons, an account of the appropriateness of blame should not. A theory of the appropriateness of blame should have something to say about the sorts of agents we might expect to see blaming and the sorts of agents we might expect to see blamed, and it should have something to say about the sorts of circumstances where we might expect blaming to take place. Consider McKenna's case again. It's true that the continued existence of the planet is important. However, this is not a concern which regularly arises (or has ever arisen) with regard to any particular instance of blaming. However, things are different with regard to the conflicts between blaming and our other potential responses, conflicts which are ubiquitous in practice. This

means that a theory of blaming should have something to say about the matters that could tell against blaming. The theory need not account in particular for all such matters, but it should account for those which arise with sufficient regularity such that an ordinary agent should possess a ready sensitivity to those sorts of facts.

5.3.2 Harris as an overlap case

We can now return to Harris's case, looking at it as an overlap case. With the historicists, I assume for purposes of argument that Harris is reasons-responsive.[138] In contrast with the historicists, however, I take it that his reasons-responsiveness is sufficient for his responsibility. And his crimes are horrific, malicious, and devastating. Therefore, I assume for purposes of argument that Harris is blameworthy for those crimes and that we as outside agents who have played no particular role in bringing about his wrongdoing have significant pro tanto reason to blame and resent Harris.

As I've argued, however, blame is not the only response we have reason to experience. As Watson and McKenna note and as former California Governor Pete Wilson noted, we should feel sympathy toward Harris. His upbringing was horrific, and so we should feel for him both for the suffering that his upbringing surely caused and for the lost opportunity to develop a healthy moral psychology. And we should resent those who mistreated Harris, both those in his family and those in

---

[138] Although they are not committed to Harris in particular being reasons-responsive, they need the bad-history agents to be reasons-responsive in order to do the needed dialectical work. In fact, I presume that Harris was in fact not fully reasons-responsive. I assume that his horrific bad-history compromised the development of his normative psychology, as I have explained previously.

institutional positions to help him. And, importantly, we should also feel sympathy toward Harris's victims--Baker, Mayeski, and their families. The innocent victims suffered terribly and without reason. So we should see Harris's case as not just one where resenting Harris is at issue but also one where these responses (and probably many others besides) are at issue as well.

The resentment we might feel toward Harris is in conflict with the other potential responses. Resentment conflicts phenomenologically with sympathy, and so it will be difficult to feel the full force of both the resentment due Harris for his culpable wrongdoings and the sympathy due him for the way he was treated in the past. And, perhaps surprisingly, it might be difficult to feel both the full force of the resentment due Harris and the full force of the sympathy due his victims for the suffering he caused.[139] Resenting Harris also requires that we see Harris as a wrongdoer, that we see him in the light of his wrongdoing. This requires that we attend to his wrongs, and that constrains the attention we can direct toward the way he was treated in his past and might even to some degree constrain the attention we can direct toward the ways his victims have suffered. For instance, someone concentrating on the suffering of Harris's victims might be fully engaged in sympathetic imagining and in identifying reparative possibilities--which might

[139] As Dana Nelkin has noted, I have left it open that the reactive attitudes might not include any particularly distinctive phenomenology. Accordingly, there might be cases where there is no phenomenological conflict. However, because I take the seeing-as element to be central and necessary to the reactive attitudes, there will be conflicts there.

make the work of paying attention to the wrongdoer more difficult (though I expect some readers will find this exercise easier than others).

Because of this conflict, we can fully blame Harris only if we forego or constrain those other responses or change ourselves so as to remove the conflict. With dutiful attention to the grounds of blameworthiness--the target agent's responsibility for a wrongdoing--the agent may experience the full degree of blame. But fully experiencing blame in the face of these psychological conflicts will usually require foregoing the full experience of the other, also-grounded response. In Harris's case, these opportunity costs are quite significant. There is good reason to resent those who mistreated Harris, to feel sympathy for Harris, and to feel sympathy for his victims. These are very important responses, and so to fail to fully experience those responses is a significant loss.

Moreover, we should be susceptible to both resentment and sympathy in just the way that leads to this conflict. Resentment is a way of acknowledging the moral significance of wrongdoing and of acknowledging the moral importance of victims of wrongdoing. This is a reaction not just to the harm suffered, but a reaction to the wrong and how it impacts our relationships as moral agents. Those of us within the Strawsonian tradition take seriously the role of resentment in marking our moral sociality. Sympathy, too, is critically important. Intuitively, sympathy is a way of acknowledging the importance of someone who has suffered, and it plays an important role in bonding and in motivating moral behavior.

Accordingly, it is inappropriate for us to feel full resentment toward Harris.[140] It may be appropriate for us to feel some resentment toward Harris, as he did commit a significant wrong. But it is also important to experience the competing responses at least in part, and that importance makes it wrong for us to fully resent Harris. So long as some mixed response is possible, it seems right to respond partially to all of the various grounds at play, the exact balancing a further question.

This assessment of Harris's case fits well with our experienced response to hearing about the case. In fact, it provides two explanatory payoffs. First, we experience a mixed response to Harris, one of both blame and something like exculpation, and the overlap argument explains this phenomenon, as I've explained. Second, the overlap argument also explains why we find the bad-history cases distinctively uncomfortable, why we find, in Watson's word, our response to Harris to be one of "ambivalence" (1987, p. 275). Harris is fully blameworthy, but we should not fully blame him. And Harris is fully sympathy-worthy, but we should not feel fully developed sympathy for him. The resulting mixed response is a second-best response, limited by our finite psychologies. The right response for us might be the best that we can do, but it leaves something important out. Our ambivalence can be seen as reflecting our sense that we are in some way falling short, even if we are doing exactly as we should.

---

[140] Alternatively, and more conservatively, it is not inappropriate not to feel resentment toward Harris. However, since the competing responses are responses due agents other than the blaming agent, I do not think those competing responses are supererogatory or merely permitted in the way that being charmed might be.

Thus there are two powerful explanations a reasons-responsiveness theorist might offer to the bad-history cases: bad history might have compromised the agent's reasons-responsiveness, thereby indirectly undermining her responsibility, and bad history might ground competing responses like sympathy, thereby undermining blame but not blameworthiness. And the more significant the bad-history, the more compelling these explanations will be. That means that the cases which might have seemed the most promising grist for the historicists--cases where the agents' bad histories make blame seem most inappropriate--are also the cases where these explanations are the most compelling. At least in the sorts of real-world bad-history cases we actually confront, the ordinary reasons-responsiveness account of moral responsibility faces no explanatory deficit requiring us to adopt historicism.

5.4 The artificial bad-history agents

My assessment of the bad-history cases turns on the complications those cases all involve. Just as the bad history leaves me reticent to fully blame Harris, it also gives me substantial reason to be skeptical that he has developed the necessary reasons-responsiveness capacities as well as substantial reason to be sympathetic toward him. Thus, while Harris's case is an importantly representative case, it might not be a particularly fit test case. Philosophers have thus designed artificial vignettes precisely to avoid just the sorts of explanations I've offered. If they succeed in that, and if they also succeed in eliciting compelling intuitions about compromised responsibility, then we will be pushed to add a historicist element to

the theory of moral responsibility. Here, I offer my skepticism that the fictional accounts can simultaneously achieve both of those goals.

Many of the fictional bad-history vignettes function by depicting an agent whose behavior (often, but not always, wrongdoing) is arguably the product of a drastic change in their personality brought about by outside intervention. Not all of the fictional bad-history vignettes have these features, but many do, as these features are a good way to prompt the audience to see that the behavior was in an important sense not the agent's own. Consider Pereboom's (2001, pp. 112–126, 2014, pp. 74–82) Professor Plum, the most famous such case:

> Professor Plum decides to murder White for the sake of some personal advantage, and succeeds in doing so. … [T]he action satisfies the reasons-responsiveness condition advocated by John Fischer and Mark Ravizza (1998): Plum's desires can be modified by, and some of them arise form, rational consideration of his reasons, and if he believed that the bad consequences for himself that would result from his killing White would be more severe than he actually expects them to be, he would not have decided to kill her. … Plum has the general ability to grasp, apply, and regulate his actions by moral reasons. (2014, p. 75)

However, Plum's killing of White can be traced to outsiders' intervention:

> A team of neuroscientists has the ability to manipulate Plum's neural states at any time by radio-like technology. In this particular case, they do so by pressing a button just before he begins to reason about his situation, which they know will produce in him a neural state that realizes a strongly egoistic reasoning process, which the neuroscientists know will deterministically result in his decision to kill White. (2014, p. 76)

Pereboom claims that we do not see Plum as responsible for killing White and that this response shows that we should accept historicism about responsibility--indeed that we should accept incompatibilism.

Or consider Al Mele's contrast between Brainwashed Beth and Evil Chuck. Evil Chuck is a deeply vicious agent who enjoys killing people, and his terrible character is his own craftsmanship. When he was younger, Chuck would torture animals despite his distaste for the act just to ensure that he could do as he pleased without regard for conventional morality. Mele then has us consider Beth:

> When Beth crawled into bed last night she was an exceptionally sweet person, as she always had been. Beth's character was such that intentionally doing anyone serious bodily harm definitely was not an option for her: her character—or collection of values—left no place for a desire to do such a thing to take root. Moreover, she was morally responsible, at least to a significant extent, for having the character she had. But Beth awakes with a desire to stalk and kill a neighbor, George. Although she had always found George unpleasant, she is very surprised by this desire. What happened is that, while Beth slept, a team of psychologists that had discovered the system of values that make Chuck tick implanted those values in Beth after erasing hers. (2013, p. 169)

Beth proceeds to act on the implanted Chuck-like values, killing George. Mele thinks that Beth is not autonomous but that Chuck is, and he thinks that these cases show that the conditions of autonomy must include a historical element.[141]

For both Beth and Plum, we are to see the agent's wrongdoing as the product of some outside factor, and the outside factor's role is to prompt us to intuit that the agents are not blameworthy. Pereboom and Mele thus think that these cases show that those agents' bad histories must be accounted for in assessing the agents. What can the ahistoricist, reasons-responsiveness theorist say about these cases? Here I offer three responses: in fact, we should see these agents as blameworthy; but if

[141] I set aside the difference between the sort of autonomy which interests Mele and the sort of responsibility which interests me.

they are not blameworthy, the ordinary, ahistoricist explanations will suffice; and, in any case, we should not trust our intuitions in these bizarre fabrications.

5.4.1 The blameworthiness of the artificial agents

Although Pereboom and Mele see the artificial bad-history agents as plainly not blameworthy, it is open to the theorist to say that both agents are deserving of blame.[142] Both cases are constructed so as to make clear that the killings are manifestations of the agents' malicious characters, and we are told that both agents have the properties sufficient for the ahistoricist, reasons-responsiveness account of moral responsibility. If I dwell on Plum's reasons-responsiveness, think about his likely awareness that the killing was wrong, and think about his insensitivity to the suffering and harm he caused, I am inclined to think that, whatever his history, Plum has become vicious, that his behavior was vicious, and that he is an appropriate target of blame. As Fischer writes, "Although Plum is manipulated by others (without his knowledge or consent), he is not forced or compelled to act as he does; thus, he is not a robot--he has a certain minimal measure of control, and moral responsibility is associated with control (of precisely this sort)" (2004, p. 157).[143] Likewise, if I focus on Beth's behavior, if I keep in the fore of my mind that, whatever the source of her ill character, Beth had the requisite reasons-responsiveness

[142] McKenna (2008b) calls this the "hard-line reply." The hard-line reply denies that the manipulated agent is not responsible. This is the line Frankfurt (not a reasons-responsiveness theorist) takes. As I quoted Frankfurt before: "The fact that someone is a pig warrants treating him like a pig" (2002, p. 28).

[143] Fischer then proceeds to distinguish responsibility for wrongdoing from blameworthiness, in large part by appealing to his taking-responsibility account. I reject that distinction and the historicism of that account, as explained previously.

capacities at the time, and if I make myself imagine someone who feels the urge to kill, who has all the rich capacities needed to control her urge, and yet indulges that urge, I am inclined to think that Beth's behavior is vicious, and that she is the appropriate target of blame. The science-fiction elements of the stories are distracting, but perhaps not impossibly so. When I focus on the agents' indulging in their inclinations to wrongdoing despite being in the possession of rich capacities to resist, I am inclined to blame both fictional agents. If this method of focusing produces similar results widely, then the fictional cases are grist for the reasons-responsiveness theory, not a challenge to it.

5.4.2 Explaining the artificial bad-history intuitions

Nonetheless, many people respond to Plum, Beth, and similar artificial cases by reporting that they find blame inappropriate. In many of the artificial cases, this is exactly the verdict the reasons-responsiveness theory should lead us to expect. As was true for the real-world bad-history agents, blame might be inappropriate because the agent is not blameworthy, and blame might inappropriate even for a blameworthy wrongdoer because of further, countervailing factors.

First, despite the professed aims of the vignettes' constructors, we should be skeptical that the artificial cases can be constructed such that we are in a position to comfortably conclude that the fictional agents are reasons-responsive. If Plum and Beth are not reasons-responsive, then the theory can readily accommodate the urged intuition that they are not appropriately to blame. Pereboom and Mele claim that the outside intervention leaves intact the features picked out by compatibilist

accounts of responsibility, e.g., the agent's reasons-responsiveness. We should be skeptical that the intervention can be so precise. This is true with respect to the capacities at issue, and it is particularly true with respect to the implanted desires. In many of the fictional cases, the desires at issue are particularly strong desires; they must be strong to guarantee the resulting unwarranted wrongs. But the presence of particularly strong desires raises the possibility that the desires are so strong as to compromise the agent's responsibility.[144] Consider Plum's desire to kill White. Did that strong desire interfere with his taking seriously the possibility of restraint, from recognizing the moral reasons which opposed the killing? Or, if the desire did not keep Plum from recognizing that he should not kill White, was it so strong he could not resist the desire? We might imagine Plum acting like an addict, driven by an urge against his own better judgment. The more I dwell on the strength of the desire, the more plausible these possibilities seem. It is not clear to me that there is room to conclude both that Plum, Beth, and the rest were not saddled with desires of this sort and that we nonetheless have imagined the relevant desire in the right sort of richness required to generate a reliable intuition about the case. This

[144] Fischer (2014, pp. 204–205) suggests a dilemma: either the desire is not strong enough to guarantee the wrongdoing, or it is strong enough to constitute a responsibility-compromising irresistible urge. I think the problem for the historicist can be made even more pressing: to prompt the right intuitions, the audience should be convinced that, but for the implanted desire, the agent would have resisted the vicious wrong. The greater the wrong, the greater the motivation we should have expected the agent to experience to resist performing the wrong. Therefore, the greater the wrong, the stronger the urge must be. If the urge is not at least as strong as the wrongdoing is wrong, then we might expect the outside intervention to do little more than to reveal the agent's prior bad character. There would then be no problem holding the agent responsible and blameworthy.

suspicion is heightened when I recall that, while almost everyone accepts that there might be desires so strong as to excuse, our understanding of such desires remains relatively undeveloped.

Second, even if Pereboom and Mele have successfully constructed cases of blameworthy artificial agents, we should suspect that these are nonetheless not agents who are properly to blame. We can explain this by appealing to the overlap assessment without needing to appeal to compromised responsibility. Both Plum and Beth are overlap cases. We have reason to resent them because of their culpable wrongdoing, but we also have reason to resent those who have manipulated them and reason to sympathize with them for having been manipulated.[145] Both have been manipulated in surreptitious fashion, leading both to commit vicious murders. We have good pro tanto reason to feel sympathy for them on these grounds. As with Harris, the various responses should compete, and so we cannot fully resent Plum and Beth without foregoing these other important responses. Because these responses conflict, the right response to both Plum and Beth is a mixed, conflicted response. Here, too, the reasons-responsiveness theorist can explain why the artificial bad-history agents are not properly to blame without any need to modify the underlying account of moral responsibility.

5.4.3 Rejecting the artificial bad-history intuitions

The bad-history cases are designed to generate intuitions which avoid these explanations. If they are successful in doing so, then these explanations, and with

[145] Michael Bratman (2000), Nomy Arpaly (2002), McKenna (2012), and others have made this point about the manipulation cases.

them my case for ahistoricism, fail. But we should be skeptical that the bad-history cases can be expected to generate the right sort of intuition.[146] The matters are too controverted, abstract, and bizarre for the resulting intuitions to be of significant probative force. This means that seeing them as noise or error is not a particularly high price.[147]

The first problem with the intuitive data is that it is not clear what the intuitions are supposed to be intuitions about. There is disagreement in the way that the intuitions are described. For Haji and Pereboom, the intuition is to regard the agent's responsibility. For Mele, the intuition is sometimes about responsibility but sometimes about free or autonomous action. For McKenna, the intuition is about the appropriateness of blame. It is not clear whether these philosophers are having intuitive judgments about different topics (and thus often talking past each other) or whether some or all of them are mistaken as to the nature of their own intuitions. This complexity is not surprising given continuing disagreements about the concepts at issue. Is there one central sort of responsibility at issue, two, or many? Is the sort of responsibility at issue the realist sort or the response-dependence sort? What exactly is the relationship between responsibility and blameworthiness? If it is

---

[146] Not everyone is so skeptical. Some think that the science-fiction cases are particularly suited to the task, since they allow us to isolate and even highlight the factors we are curious to investigation. See, for example, Pereboom (2014, p. 95 quoting Nelkin from conference comments approvingly) and Carolina Sartorio (2016, p. 165).

[147] Because there are ostensibly conflicting intuitions--ordinary intuitions of responsibility about ordinary agents and fantastic intuitions of non-responsibility about artificial agents--many philosophers will categorize one or the other class of intuitions as in error. Thus we see Pereboom (2014, p. 88) offering an argument that we should expect our intuitions about ordinary cases to be in error.

not settled among philosophers what exactly "responsibility" is to pick out, it is not clear how I could be confident that my intuition regards that phenomenon, whatever it is supposed to be.

The second problem is that we cannot expect reliable data from the artificial cases. Assume that there is some agreed sense of realist, accountability responsibility tied to blameworthiness (to pick one sense of responsibility at issue), distinct from attributability responsibility, and distinct from blame's being all-things-considered appropriate. We would need to be confident that our intuitions are properly responding to that phenomenon in particular. Ishtiyaque Haji (2013), for example, acknowledges that the historicist must take care to elicit an intuitive response which avoids competing explanations. He insists that we are able to keep these competing features at bay, to focus our moral imaginations on the cases in an intentional way, and to thereby generate an intuition which we can be confident is not responding to those competing features. But I cannot focus my intuitions like that. When I attempt to introspectively examine my response to a case and thereby focus on blameworthiness in particular, I lose my grasp upon the case and my response. Of course, that could be merely my own shortcoming. But I have little beyond Haji's own assertion that his intuitions are significantly more sensitive, nuanced, and controllable than my own.

That our intuitions in these cases might not be reliable should not be surprising. The cases are strange, and strange cases might be expected to produce strange data. We see this worry from King: "the intuition here regards a

significantly fringe case of action, involving a mechanism that is, at present, technologically impossible. One might reasonably think that our intuitions in such cases are defeasible" (2013, p. 69). We see a similar worry from McKenna:

> Our intuitions have evolved along with our ordinary practices. It is only to be expected that when those intuitions are tested in extremely different contexts, contexts which differ radically from the ones out of which they evolved, they will be indecisive. … [W]hen we test our intuitions against wildly divergent contexts, we are certainly not licensed to draw decisive conclusions. We are not sure what to make of them. (2008a, p. 157)

Granting that intuition-pumping is in general a reliable method for getting valuable data about normative facts, we might nonetheless have particular reason to be suspicious of the method in such bizarre cases.

Although I do not here offer a full defense of this suspicion, I can offer several reasons in support of my skepticism. First, the bizarre features will be particularly attention-getting, and so they might distort our verdicts by distracting us from more quotidian but morally relevant features. A science-fiction manipulation might capture our attention just by being so different from ordinary experience, and this might distract us from the preserved elements of the agent's psychology. I focus on Plum's manipulation, rather than focusing on the state of his capacities, the nature of his control over his behavior, or even his wrongdoing.

Second, even if we do keep all of the elements of the vignette duly in mind, why assume that we can readily and accurately imagine whatever is presented in the text of the vignette? I know that I am to imagine that Plum has been manipulated in a way that is both preserving of the ahistoricist conditions and efficacious--but I

don't know what it is to imagine that. Chandra Sripada registers a similar complaint: "The philosophers who have advanced manipulation cases actually provide precious little information to help…, and instead make only broad references to 'rigorous practices of conditioning', vaguely Skinnerian kinds of behavioral engineering, and 'scripting' major life events" (2012, p. 569 citations omitted). This is not the sort of language that aids the imaginative process you might expect to prompt a reliable intuition. The abstractions do little to help illuminate the fringe which worried King.

Third, we should expect our ordinary experiences to infect our intuitions about extraordinary experiences.[148] Consider the 'ordinary' manipulations familiar to us, such as hypnosis or drugs, things which render the agent readily suggestible. These leave us open to manipulation, but they seem to do so in ways which would provide good explanations of the manipulation intuition for the compatibilist. It is more than plausible that drugs and hypnosis compromise reasons-responsiveness. And so our experiences with such cases (real or imagined) might make us susceptible to judging other manipulated agents non-responsible, even if there is not specific evidence of compromised reasons-responsiveness.

This skepticism about the probative value of the intuitions prompted by the artificial cases is supported by the empirical data. Dylan Murray and Eddy Nahmias (2014) report that the explanation of an agent's moral responsibility and free will judgments appears sensitive to the perception (vignette-induced or otherwise) of an agent's practical reasoning mechanisms being bypassed--not determinism. This was

[148] Pereboom raises a similar worry about ordinary experience in trying to defuse the ordinary intuitions of responsibility.

the case even though the involved vignettes were not designed to prompt bypassing judgments. That means that at least some participants were (perhaps against the vignettes) inferring that causal determinism worked by bypassing the relevant agential structures, robbing the intuition data of its probative effects with respect to the direct relevance of history to moral responsibility.

Sripada (2012) gives a similar indictment of the probative value of the artificial cases. He notes that people may imagine cases differently, and they may respond to different features of the cases (such that we cannot be sure that all of the elements of the vignette are playing their intended role). There was a high correlation in the data he examined between participants assessing that an agent lacked free will and participants judging there to be corruption in either the information available to the agent or the agent's moral psychology. Importantly, "reading the Manipulation vignette … had no significant effect at all on free will ratings over and above the effect it had on producing judgments that the agent suffered from certain kinds of psychological damage" (Sripada, 2012, p. 582). The shadows cast by manipulation upon the vignette agent's contemporary psychology fully explained the free-will judgments. As Sripada concludes, "the results of the studies reported in this paper suggest that these extra historical conditions are actually unnecessary" (2012, p. 567 n.2). These reports suggest that it will be difficult to craft a vignette which prompts the needed intuition and also that the actual data supports the ahistoricist against the historicist.

Given the complexity of the concepts at issue, the vagueness of the manipulation at issue, and the empirical data, I am skeptical that the bad-history cases will readily provide the particular intuition which the historicists need. I am also pessimistic that the historicists will be able to design a case which might more clearly elicit an intuition about blameworthiness without suggesting the competing explanation from the overlap cases. Any appeal to a bad history will, by the very fact of the badness of the history, invite the explanation from the overlap cases. And even if some delicately constructed case could be identified, the case would have to be imagined precisely, and the intuition would have to regard the precise normative facts. But the complications of the case should undermine our confidence that we've imagined the case correctly, holding the right features in mind and avoiding unnoticed interference. And to ensure that we've imagined the case clearly, we might need to simplify the case, but then it will be difficult to be certain that our intuitive responses are about the right normative facts, about blameworthiness as opposed to the appropriateness of blame. And so I am inclined to think that it is more likely that Haji is mistaken about his own intuitions than that he is capable of generating the sort of intuitions needed to preserve the historicists' argument. But this, of course, remains open.

Thus I take my explanations to defuse the most popular argument for historicism, at least given the cases so far offered to support the argument. I can offer satisfying explanations of both real-life cases like Robert Alton Harris and artificial cases like Professor Plum. This does not mean that historicism is false; it

means only that the most popular argument for historicism is unsound. Other arguments for historicism--e.g., the consequence argument made popular by Peter van Inwagen (1983)--are left untouched. However, while my argument is consistent with the truth of historicism, it can play a key role in the defense of ahistoricism, by defusing one of the central challenges ahistoricism faces.

Chapter 5, in part, is currently being prepared for submission for publication as Craig Agule, "Being Sympathetic to Bad-History Wrongdoers." The dissertation author was the sole investigator and author of this paper.

Conclusion

This concludes a powerful case for an ahistorical, reasons-responsiveness account of moral responsibility. On my core account, responsibility has to do with the sort of agent you are when you act. You are responsible for your behavior if and only if your behavior manifests your quality of will, and your behavior manifests your quality of will if and only if you act while in the possession of central, agential capacities.

Reasons-responsiveness accounts of moral responsibility are widely popular, but most advocates augment the core account by adding historical elements of various sorts. The historical elements are most often added because the core account is seen as unable to explain our intuitive responses to important problem cases. Accordingly, I defend the ahistorical, reasons-responsiveness account by taking on those important problem cases, by showing that the core reasons-responsiveness account can provide satisfying explanations for those cases without requiring any historical commitments.

I have shown that the ahistorical, reasons-responsiveness account can provide satisfying explanations of the cases often thought to compel historicism. However, that those cases do not give us reason to accept historicism does not mean that the historicist might not find support for historicism elsewhere. For example, historicism might yet be grounded in appeal to first principles about fairness or in further explorations of the notion of control. Although the devil is in the details, there is no reason to think that my explanations of the potentially troublesome

cases would be inconsistent with such historicist accounts of moral responsibility, and so I have not begged the question against the historicist. All that said, while my arguments do not provide a final case against historicism, they do give good reason to accept an ahistorical account of moral responsibility.

This means that we have good reason to think that moral responsibility is an ahistorical phenomenon. Moral responsibility has to do with the sort of agent you are, not with what has happened to you in your past. The past is causally and epistemically relevant to moral responsibility, but moral responsibility is essentially ahistorical. Seeing this is important for getting the right theory of moral responsibility, for getting our moral responsibility verdicts correct, and for understanding the sort of control implicated in moral responsibility.

Although my argument provides a compelling answer to the question of historicism, it also pushes us toward a number of further questions. Some of these questions arise out of the explanations offered by the ahistoricist account for the tracing and bad-history cases. For instance, once we recognize that we should account for the tracing cases by focusing on blameworthy acts of self-incapacitation, we need to consider whether those acts are more widespread than we might have originally noticed. Are we overlooking a tremendous number of significant wrongdoings just because they do not result in an additional, compromised wrongdoing? Likewise, the explanation of the bad-history cases appeals to the psychological conflict between resentment and sympathy in those cases. Recognizing that conflict can help us explain our intuitive responses to those cases,

but it should also push us to ask whether that conflict is fixed and to ask whether, in light of that conflict, we might better structure our conflicting responses. And, of course, while it is widely accepted that an agent's history is causally and epistemically relevant to his contemporary reasons-responsiveness, the exact nature of this relevance, and in particular the exact nature of the causal development of the capacities, needs illuminating.

More broadly, my arguments point to broader questions about the nature of the reasons-responsiveness account. For instance, the argument about the bad-history cases focuses on the relationship between the fittingness of the reactive attitudes and the reasons we might have to experience them (or not)--a relationship that is widely assumed but underexplored. The reasons-responsiveness account of moral responsibility explains the conditions of the reactive attitudes being fitting. Does fittingness then ground a practical reason of some kind to experience the reactive attitudes? What sort of reason would this be? If this reason derives from the role that the reactive attitudes play in the good life, what exactly is the nature of that role, and how do other concerns interact with that role? Similarly, my arguments put pressure on the notion of capacity at the heart of the reasons-responsiveness account. Both the general notion of a psychological capacity and the particular reasons-responsiveness capacities are intuitively attractive, but the working of the capacities remains underexplored. These capacities will not admit of a standard conditional analysis--for example, you have the capacity if you would perform if you tried to perform--as the capacity to try is at the heart of the issue. So my arguments

push us to identify the relevant notion of capacities that could do the work instead. If we had that fully developed notion of capacity, we could see how it relates to quality of will, how it admits of degrees, and the like.

That my arguments for ahistoricism raise these further questions is no particular weakness of the account--these are questions which all reasons-responsiveness theorists should face, and my arguments merely help to bring them into focus. So my arguments here both establish a compelling case for an ahistoricist, reasons-responsiveness account of moral responsibility and provide a framework for further investigations into moral responsibility.

WORKS CITED

Agule, C. (n.d.). Paying Attention to Standing.

Agule, C. (2016). Resisting Tracing's Siren Song. *Journal of Ethics & Social Philosophy*, *10*(1), 1–24.

Alexander, L. (2013). Causing the Conditions of One's Defense: A Theoretical Non-Problem. *Criminal Law and Philosophy*, *7*(3), 623–628.

Allais, L. (2008a). Dissolving Reactive Attitudes: Forgiving and Understanding. *South African Journal of Philosophy*, *27*(3), 179–201.

Allais, L. (2008b). Wiping the Slate Clean: The Heart of Forgiveness. *Philosophy & Public Affairs*, *36*(1), 33–68.

Aristotle. (1998). *The Nicomachean Ethics*. (W. D. Ross, Trans.). Oxford University Press.

Arpaly, N. (2002). *Unprincipled Virtue*. Oxford University Press.

Bell, M. (2013). The Standing to Blame: a Critique. In D. J. Coates & N. A. Tognazzini (Eds.), *Blame: Its Nature and Norms* (pp. 263–281). Oxford University Press.

Bratman, M. E. (2000). Fischer and Ravizza on Moral Responsibility and History. *Philosophy and Phenomenological Research*, *61*(2), 453–458.

Brink, D. O. (2004). Immaturity, Normative Competence, and Juvenile Transfer: How (Not) to Punish Minors for Major Crimes. *Texas Law Review*, *82*, 1555–1585.

Brink, D. O. (2013). Responsibility, Incompetence, and Psychopathy. *The Lindley Lectures*, (2013).

Brink, D. O., & Nelkin, D. K. (2013). Fairness and the Architecture of Responsibility. *Oxford Studies in Agency and Responsibility*, *1*, 284–313.

Christman, J. (1991). Autonomy and Personal History. *Canadian Journal of Philosophy*, *21*(1), 1–24.

Clarke, R. (1993). Toward a Credible Agent-Causal Account of Free Will. *Noûs*, *27*(2), 191–203.

Coates, D. J., & Swenson, P. (2013). Reasons-Responsiveness and Degrees of Responsibility. *Philosophical Studies*, *165*(2), 629–645.

D'Arms, J., & Jacobson, D. (2000). The Moralistic Fallacy: On the "Appropriateness" of Emotions. *Philosophical and Phenomenological Research*, *61*(1), 65–90.

Dennett, D. C. (1984). *Elbow Room: The Varieties of Free Will Worth Wanting*. MIT Press.

De Sanctis, V. A., Nomura, Y., Newcorn, J. H., & Halperin, J. M. (2012). Childhood Maltreatment and Conduct Disorder. *Child Abuse & Neglect*, *36*(11), 782–789.

Dijksterhuis, A., & Aarts, H. (2010). Goals, Attention, and (un) Consciousness. *Annual Review of Psychology*, *61*, 467–490.

Dworkin, R. (1986). *Law's Empire*. Harvard University Press.

Fine, C., & Kennett, J. (2004). Mental Impairment, Moral Understanding and Criminal Responsibility. *International Journal of Law and Psychiatry*, *27*(5), 425–443.

Fine, K. (1994). Essence and Modality. *Philosophical Perspectives*, *8*, 1–16.

Fischer, J. M. (2000). Responsibility, History and Manipulation. *The Journal of Ethics*, *4*(4), 385–391.

Fischer, J. M. (2004). Responsibility and Manipulation. *The Journal of Ethics*, *8*(2), 145–177.

Fischer, J. M. (2012). *Deep Control: Essays on Free Will and Value*. Oxford University Press.

Fischer, J. M. (2014). Review of Free Will, Agency, and Meaning in Life, by Derk Pereboom. *Science, Religion & Culture*, *1*, 202–208.

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

Fischer, J. M., & Tognazzini, N. A. (2009). The Truth About Tracing. *Noûs*, *43*(3), 531–556.

Fischer, J. M., & Tognazzini, N. A. (2011a). The Physiognomy of Responsibility. *Philosophy and Phenomenological Research*, *82*(2), 381–417.

Fischer, J. M., & Tognazzini, N. A. (2011b). The Triumph of Tracing. In *Deep Control: Essays on Free Will and Value* (pp. 206–234). Oxford University Press.

Foot, P. (2001). *Natural Goodness*. Oxford University Press.

Frankfurt, H. G. (1969). Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy*, *66*(23), 829–839.

Frankfurt, H. G. (1971). Freedom of the Will and the Concept of a Person. *Journal of Philosophy*, *68*(1), 5–20.

Frankfurt, H. G. (2002). Reply to John Martin Fischer. In S. Buss & L. Overton (Eds.), *Contours of Agency: Essays on Themes from Harry Frankfurt* (pp. 27–32). MIT Press.

Fricker, M. (2016). What's the Point of Blame? A Paradigm Based Explanation. *Noûs*, *50*(1), 165–183.

Haji, I. (2000). On Responsibility, History and Taking Responsibility. *The Journal of Ethics*, *4*(4), 392–400.

Haji, I. (2010). Psychopathy, Ethical Perception, and Moral Culpability. *Neuroethics*, *3*(2), 135–150.

Haji, I. (2013). Historicism, Non-historicism, or a Mix? *The Journal of Ethics*, *17*(3), 185–204.

Hart, H. L. A. (1968). *Punishment and Responsibility: Essays in the Philosophy of Law*. Oxford University Press.

Heim, C., & Nemeroff, C. B. (2001). The Role of Childhood Trauma in the Neurobiology of Mood and Anxiety Disorders. *Biological Psychiatry*, *49*(12), 1023–1039.

Hieronymi, P. (2004). The Force and Fairness of Blame. *Philosophical Perspectives*, *18*(1), 115–148.

Hume, D. (2000). *A Treatise of Human Nature*. Oxford University Press.

Hurley, E. A., & Macnamara, C. (2010). Beyond Belief: Toward a Theory of the Reactive Attitudes. *Philosophical Papers*, *39*(3), 373–399.

Husak, D. (2010). The De Minimis "Defence" to Criminal Liability. In *The Philosophy of Criminal Law* (pp. 362–392). Oxford University Press.

Husak, D. (2012). Intoxication and Culpability. *Criminal Law and Philosophy*, *6*(3), 363–379.

Judisch, N. (2005). Responsibility, Manipulation and Ownership. *Philosophical Explorations*, *8*(2), 115–130.

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

Khoury, A. C. (2012). Responsibility, Tracing, and Consequences. *Canadian Journal of Philosophy*, *42*(3), 187–207.

King, M. (2013). The Problem with Manipulation. *Ethics*, *124*(1), 65–83.

King, M. (2014). Traction without Tracing: A (Partial) Solution for Control-Based Accounts of Moral Responsibility. *European Journal of Philosophy*, *22*(3), 463–482.

King, M. (2015). Manipulation Arguments and the Moral Standing to Blame. *Journal of Ethics & Social Philosophy*, *9*(1), 1–20.

Levy, N. (2007a). Norms, Conventions, and Psychopaths. *Philosophy, Psychiatry, & Psychology*, *14*(2), 163–170.

Levy, N. (2007b). The Responsibility of the Psychopath Revisited. *Philosophy, Psychiatry, & Psychology*, *14*(2), 129–138.

Macnamara, C. (2015). Reactive Attitudes as Communicative Entities. *Philosophy and Phenomenological Research*, *90*(3), 546–569.

Matheson, B. (2014). Compatibilism and Personal Identity. *Philosophical Studies*, *170*(2), 317–334.

McKenna, M. S. (1998). The Limits of Evil and the Role of Moral Address: A Defense of Strawsonian Compatibilism. *The Journal of Ethics*, *2*(2), 123–142.

McKenna, M. S. (2008a). A Hard-line Reply to Pereboom's Four-Case Manipulation Argument. *Philosophy and Phenomenological Research*, *77*(1), 142–159.

McKenna, M. S. (2008b). Putting the Lie on the Control Condition for Moral Responsibility. *Philosophical Studies*, *139*(1), 29–37.

McKenna, M. S. (2012). *Conversation and Responsibility*. Oxford University Press.

Mele, A. R. (1990). Irresistible Desires. *Noûs*, *24*(3), 455–472.

Mele, A. R. (2013). Manipulation, Moral Responsibility, and Bullet Biting. *The Journal of Ethics*, *17*(3), 167–184.

Menges, L. (2017). The Emotion Account of Blame. *Philosophical Studies*, *174*(1), 257–273.

Milam, P.-E. (2014). Abolitionism and the Value of the Reactive Attitudes. *eScholarship*. Retrieved from http://escholarship.org/uc/item/78p3g0vc

Milam, P.-E. (2016). Reactive Attitudes and Personal Relationships. *Canadian Journal of Philosophy*, *46*(1), 102–122.

Morse, S. J. (2002). Uncontrollable Urges and Irrational People. *Virginia Law Review*, *88*(5), 1025–1078.

Murray, D., & Nahmias, E. (2014). Explaining Away Incompatibilist Intuitions. *Philosophy and Phenomenological Research*, *88*(2), 434–467.

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

Nelkin, D. K. (2014). Moral Responsibility, Conversation, and Desert. *Philosophical Studies*, *171*(1), 63–72.

Nelkin, D. K. (2015). Psychopaths, Incorrigible Racists, and the Faces of Responsibility. *Ethics*, *125*(2), 357–390.

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

Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.

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

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

Perry, B. D., Pollard, R. A., Blaicley, T. L., Baker, W. L., & Vigilante, D. (1995). Childhood Trauma, the Neurobiology of Adaptation, and "Use-dependent" Development of the Brain: How "States" Become "Traits." *Infant Mental Health Journal*, *16*(4), 271–291.

Plantinga, A. (1974). *The Nature of Necessity*. Oxford University Press.

Pomorski, S. (1997). On Multiculturalism, Concepts of Crime, and the De Minimis Defense. *BYU Law Review*, *1997*, 51.

Pynoos, R. S., Steinberg, A. M., & Piacentini, J. C. (1999). A Developmental Psychopathology Model of Childhood Traumatic Stress and Intersection with Anxiety Disorders. *Biological Psychiatry*, *46*(11), 1542–1554.

Rawls, J. (1955). Two Concepts of Rules. *The Philosophical Review*, *64*(1), 3–32.

Reavis, J. A., Looman, J., Franco, K. A., & Rojas, B. (2013). Adverse Childhood Experiences and Adult Criminality: How Long Must We Live before We Possess Our Own Lives? *The Permanente Journal*, *17*(2), 44–48.

Ryberg, J. (2014). Responsibility and Capacities: A Note on the Proportionality Assumption. *Analysis*, *74*(3), 393–397.

Sarch, A. F. (2015). Knowledge, Recklessness and the Connection Requirement Between Actus Reus and Mens Rea. *Penn State Law Review*, *120*(1).

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

Schelling, T. C. (1980). *The Strategy of Conflict*. Harvard University Press.

Shabo, S. (2005). Fischer and Ravizza on History and Ownership. *Philosophical Explorations*, *8*(2), 103–114.

Shabo, S. (2015). More Trouble with Tracing. *Erkenntnis*, *80*(5), 987–1011.

Sher, G. (2009). *Who Knew?: Responsibility Without Awareness*. Oxford University Press.

Shoemaker, D. (2015). *Responsibility from the Margins*. Oxford University Press.

Smart, J. J. C. (1961). Free-will, Praise and Blame. *Mind*, *70*(279), 291–306.

Smith, A. (2010). *The Theory of Moral Sentiments*. Penguin.

Smith, A. M. (2008). Control, Responsibility, and Moral Assessment. *Philosophical Studies*, *138*(3), 367–392.

Smith, H. (1983). Culpable Ignorance. *The Philosophical Review*, *92*(4), 543–571.

Sripada, C. S. (2012). What Makes a Manipulated Agent Unfree? *Philosophy and Phenomenological Research*, *85*(3), 563–593.

Strawson, P. F. (2008). Freedom and Resentment. In *Freedom and Resentment and Other Essays* (pp. 1–28). Routledge.

Stueber, K. (2017). Empathy. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*.

Stump, E. (2002). Control and Causal Determinism. In S. Buss & L. Overton (Eds.), *Contours of Agency: Essays on Themes from Harry Frankfurt* (pp. 33–60). MIT Press.

Timpe, K. (2011). Tracing and the Epistemic Condition on Moral Responsibility. *The Modern Schoolman*, *88*, 5–28.

Todd, P. (2016). Strawson, Moral Responsibility, and the "Order of Explanation": An Intervention. *Ethics*, *127*(1), 208–240.

Tognazzini, N. A. (2013). Blameworthiness and the Affective Account of Blame. *Philosophia*, *41*, 1299–1312.

Tognazzini, N. A., & Coates, D. J. (2016). Blame. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.

Van Inwagen, P. (1975). The Incompatibility of Free Will and Determinism. *Philosophical Studies*, *27*(3), 185–199.

Van Inwagen, P. (1983). *An Essay on Free Will*. Oxford University Press.

Vargas, M. (2005a). The Revisionist's Guide to Responsibility. *Philosophical Studies*, *125*(3), 399–429.

Vargas, M. (2005b). The Trouble with Tracing. *Midwest Studies in Philosophy*, *29*(1), 269–291.

Vargas, M. (2006). On the Importance of History for Responsible Agency. *Philosophical Studies*, *127*(3), 351–382.

Vargas, M. (2013). *Building Better Beings: A Theory of Moral Responsibility*. Oxford University Press.

Vuoso, G. (1987). Background, Responsibility, and Excuse. *The Yale Law Journal*, *96*(7), 1661–1686.

Wallace, R. J. (1994). *Responsibility and the Moral Sentiments*. Harvard University Press.

Wallace, R. J. (2011). Dispassionate Opprobrium: On Blame and the Reactive Sentiments. In R. J. Wallace, R. Kumar, & S. Freeman (Eds.), *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon* (pp. 348–372). Oxford University Press.

Watson, G. (1987). Responsibility and the Limits of Evil. In F. D. Schoeman (Ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology* (pp. 256–286). Cambridge University Press.

Watson, G. (1996). Two Faces of Responsibility. *Philosophical Topics*, *24*(2), 227–248.

Williams, B. (1982). *Moral Luck: Philosophical Papers 1973-1980*. Cambridge University Press.

Wolf, S. (1990). *Freedom Within Reason*. Oxford University Press.

Zimmerman, D. (2001). Thinking With Your Hypothalamus: Reflections on a Cognitive Role for the Reactive Emotions. *Philosophy and Phenomenological Research*, *63*(3), 521–541.

Zimmerman, D. (2002). Reasons-responsiveness and Ownership-of-agency: Fischer and Ravizza's Historicist Theory of Responsibility. *The Journal of Ethics*, *6*(3), 199–234.

Zimmerman, M. J. (2002). Taking Luck Seriously. *The Journal of Philosophy*, *99*(11), 553–576.