Responsibility, Control, and the Nonvoluntary

Kyle G. (Kyle Gregory) Fritz
RESPONSIBILITY, CONTROL, AND THE NONVOLUNTARY

By

KYLE G. FRITZ

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

2016
Kyle Fritz defended this dissertation on June 17, 2016.
The members of the supervisory committee were:

Randolph Clarke  
Professor Directing Dissertation

Martin Kavka  
University Representative

Alfred R. Mele  
Committee Member

Piers Rawling  
Committee Member

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

I have had the privilege and fortune of being educated by some remarkable philosophers in my life, including Michael Bishop, Jo Ellen Jacobs, Michael Jubien, Stephen Kearns, Kirk Ludwig, David McNaughton, Alfred Mele, Robert Money, Marina Oshana, Piers Rawling, and Gene Witmer. I am grateful for their passion and encouragement over the years. Thanks also to Martin Kavka for his willingness to serve as an outside member on my dissertation committee, and to Angela Smith for patiently and graciously answering a number of questions about her view.

I owe the largest debt to Randy Clarke for all his helpful feedback, careful criticisms, and thoughtful questions on various drafts of dissertation chapters and papers. My work is far clearer for Randy’s advice and attention to detail, and I will strive to be as clear and meticulous as Randy in my future writing.

In my graduate career, I have learned a great deal from several colleagues who I am fortunate enough to call friends, including Brian Coffey, Mary Marcous, Steve McFarlane, and Josh Shepherd. Gabriel de Marco, in particular, has been a great friend, and I value the conversations we’ve shared over the years. I am especially indebted to Dan Miller, who was always willing to talk about philosophical problems with me or read drafts of papers and offer comments. Dan’s feedback and friendship has been indispensible in writing this dissertation.

I wouldn’t be where I am today without support from my partner, Casey. Her encouragement and patience have sustained me through difficult times, and her friendship has made celebrating victories richer.

My friend Zane provided more financial support and distraction through my graduate career than I care to recall, and I am incredibly grateful for it. His friendship has kept me as well-rounded and grounded as a philosopher could reasonably expect.

Finally, I owe a huge debt to my parents, Greg and Deb, for their support, encouragement, and love over the years.
# TABLE OF CONTENTS

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

1. RESPONSIBILITY FOR THE NONVOLUNTARY .................................................................1
   1.1 Tension: Theory vs. Practice ............................................................................................1
   1.2 Ways to Relieve the Tension .........................................................................................3
   1.3 Kinds of Responsibility ..............................................................................................7
   1.4 The Road Ahead .......................................................................................................11

2. THE VOLITIONIST THESIS AND VOLUNTARY CONTROL ...........................................13
   2.1 Introduction ................................................................................................................13
   2.2 The Traditional Volitionist Thesis ..............................................................................13
   2.3 An Account of Voluntary Control .............................................................................19
   2.4 Choosings and Acts of Will .......................................................................................23
   2.5 Starting Over: A New Statement of the Volitionist Thesis .........................................27

3. THE VOLITIONIST THESIS AND OMISSIONS .............................................................33
   3.1 Introduction ................................................................................................................33
   3.2 Expressing Who We Are As Agents ..........................................................................34
   3.3 Exercising Control .....................................................................................................36
   3.4 The Causal Theory of Action ....................................................................................45
   3.5 Capacity Views ..........................................................................................................51

4. VOLITIONISM AND DERIVATIVE RESPONSIBILITY ..................................................55
   4.1 Introduction ................................................................................................................55
   4.2 Derivative Responsibility and Tracing ........................................................................56
   4.3 Actual Foresight .........................................................................................................58
   4.4 Could Have Foreseen ...............................................................................................63
   4.5 Should Have Foreseen ..............................................................................................64
   4.6 Taking Stock .............................................................................................................71

5. A NONVOLITIONIST ALTERNATIVE: THE RATIONAL RELATIONS VIEW .............76
   5.1 Introduction ................................................................................................................76
   5.2 The Rational Relations View ....................................................................................77
   5.3 Objections to the Rational Relations View ...............................................................81
      5.3.1 Too Close to Volitionism .....................................................................................81
      5.3.2 Too Weak to Justify Robust Moral Responsibility .............................................85
      5.3.3 Too Wide ...........................................................................................................91
      5.3.4 Too Narrow .......................................................................................................96
ABSTRACT

It seems well accepted that agents can only be morally responsible for what they voluntarily control. Yet as a matter of practice, we seem to hold agents responsible for items outside of their voluntary control, blaming others for failing to notice certain features of their environment, for having certain attitudes, or for forgetting to fulfill a promise. There is an apparent tension between philosophical theory and everyday practice.

One way to explain away this tension is by appealing to a distinction between direct and derivative responsibility. When an agent $S$ is responsible for something, $X$, in virtue of being responsible for something else that does not include $X$, $Y$, then $S$ is derivatively responsible for $X$. When $S$ is responsible for $X$ but not solely in virtue of being responsible for some other thing, $Y$, $S$ is directly responsible for $X$. One might argue that to be directly responsible for some item, one must have voluntary control over it. We are, at best, derivatively responsible for items over which we lack voluntary control, such as nonvoluntary attitudes. If we can trace back to a point at which an agent had voluntary control over some behavior and was aware of the potential nonvoluntary item that might result, the agent could be derivatively responsible for that item.

Theorists who opt to relieve the tension in this way are called volitionists. They insist that voluntary control is required for direct responsibility. Recently, however, several theorists, called nonvolitionists, have denied that voluntary control is required for direct responsibility. Instead, they claim, agents can be directly responsible for nonvoluntary attitudes and other items over which they lack voluntary control.

In this dissertation, I begin to assess the motivation for and merits of such nonvolitionist views. In chapter 2, I attempt to precisely articulate a volitionist thesis, which is more difficult than volitionists and nonvolitionists have yet appreciated. I argue that there are substantial problems with understanding the volitionist thesis as the claim that one is directly responsible only for what is under one’s voluntary control. Rather than formulating the volitionist thesis in this way, as others have done, I suggest that it is clearer and more useful to understand the volitionist thesis as the claim that agents are directly responsible only for their intentional actions.

The volitionist thesis I offer in chapter 2 does not include direct responsibility for many omissions. While volitionists might want to amend this volitionist thesis to include these
omissions, I argue in chapter 3 that they cannot do so in a justified way. If volitionists cannot comfortably extend direct responsibility to include omissions, they must account for these items derivatively via some kind of tracing strategy. In chapter 4, I argue that there are problems with proposed epistemic conditions on derivative responsibility, and some nonvoluntary faults—especially negligence and certain omissions—will prove quite difficult, if not impossible, for the tracing strategy. Accounting for responsibility for some nonvoluntary faults either directly or derivatively might be much more difficult for volitionists than they have previously appreciated. In light of this difficulty in matching our everyday responsibility ascriptions and practices, we should entertain nonvolitionist accounts.

The second half of the dissertation explores the plausibility of nonvolitionism. In chapter 5, I explain and defend one of the current leading nonvolitionist accounts: Angela Smith’s rational relations view (RRV). RRV holds that agents can be directly responsible for their attitudes in virtue of a rational connection between such attitudes and the judgments and values that underlie them. The second half of chapter 5 as well as chapter 6 are devoted to defending RRV against various objections, including a recent criticism from Neil Levy. Levy argues that direct responsibility requires consciousness of the moral significance of one’s behavior. I argue that Levy fails to establish such consciousness as necessary for direct responsibility, and then offer some revisions to the rational relations view that might help to mitigate Levy’s worries.

This modified version of RRV, which I call the rational control view (RCV), is thus poised to be a viable alternative to volitionist views. In chapter 7, I explain and defend RCV. My conclusion is modest: RCV is a serious and legitimate contender to volitionist views in light of the problems I highlight in the first several chapters. As such, RCV is worth developing further as we search for the best account of how and why agents are morally responsible for not only intentional actions, but also omissions and attitudes.
CHAPTER 1

RESPONSIBILITY FOR THE NONVOLUNTARY

1.1 Tension: Theory vs. Practice

As human agents, blame is deeply entrenched in our lives. We blame others for a variety of intentional actions and omissions: for choosing to say something spiteful, for pushing someone, for standing by and watching someone struggle to carry a heavy load rather than open the door. We typically think that this blame is justified only if the agent meets certain conditions.\(^1\) For example, it seems widely accepted that agents cannot be responsible for what is outside of their voluntary control. An agent who does not have the requisite control cannot be held responsible for her behavior, as her behavior does not express her agency in a way that can ground blame. Intentional actions and omissions, however, involve an exercise of this control such that the agent does express her agency and so can legitimately be held responsible.

Our everyday responsibility ascriptions and practices, however, seem in tension with this general principle that responsibility requires voluntary control. We blame people for a variety of faults that are not intentional actions or omissions. We blame someone for forgetting to pick us up from the airport despite a promise to be there. We blame agents for having racist or sexist attitudes. We blame a significant other of several years for failing to notice that we dislike some food. What sets these sorts of faults apart from the intentional actions and omissions, it seems, is that they are not voluntary or not under our voluntary control. One doesn’t choose to forget to pick someone up from the airport, nor does one (generally) choose to have certain attitudes. This is not to say that these sorts of faults are involuntary, or against our will in some way. Instead, they are nonvoluntary. When these items are wrong or bad, they are faults. Thus, we might refer collectively to objectionable attitudes, as well as our forgettings, negligence, failures to notice, and unintentional omissions, as nonvoluntary faults.\(^2\)

---

\(^1\) I assume here and throughout that some agents are, in fact, morally responsible creatures. Following the practice of many philosophers, I will focus largely on blame rather than praise.

\(^2\) While there is a counterpart here of, perhaps, nonvoluntary merits, I will focus almost exclusively on faults. I am uncertain that the conditions for blameworthiness neatly parallel the conditions for praiseworthiness, so I must leave questions about responsibility for nonvoluntary merits aside.
Consider the following cases of nonvoluntary faults:

*Sore Tooth:* Fiona, a politically liberal person who claims that the color of a person’s skin does not affect a person’s worth, has a toothache and seeks a recommendation for a dentist from a friend. Upon entering the dentist’s office and meeting her new dentist, Fiona feels a sudden pang of dismay when she realizes the new dentist is an African-American. (adapted from H. Smith 2011: 140)

*Fellowship:* Dan and John have both applied to the same fellowship. While both applicants have worked hard in their careers, John has put in more work than Dan has and is more deserving of the fellowship. When it is announced that John has won the fellowship, Dan feels resentment well up inside of him, which manifests in a scowl.

*Birthday:* Angela promised her close friend Dave that she would call him on his birthday. Despite her promise, Dave’s birthday comes and goes, and Angela forgets to call. (adapted from A. Smith 2005: 236)

*Bad Joke:* Ryland is very self-absorbed. Though not malicious, she is oblivious to the impact that her behavior will have on others. Consequently, she is bewildered and a bit hurt when her rambling anecdote about a childless couple, a handicapped person, and a financial failure is not well received by an audience that includes a childless couple, a handicapped person, and a financial failure. (from Sher 2009: 28)

Though these cases are underdescribed, one’s initial judgment might be that blame toward Fiona, Dan, Angela, and Ryland is deserved or fitting, even though their faults are nonvoluntary. Each of them is blameworthy for some fault that is not voluntary, but we must be clear about precisely what that fault might be.

*Sore Tooth* involves an agent having an objectionable attitude, even though the attitude is spontaneous and clashes with other attitudes she holds. Nevertheless, one might think Fiona’s

---

3 Even if one does not judge that the agent is blameworthy in every one of the above cases, one might judge that at least one of the agents is blameworthy. This judgment should be sufficient to generate the puzzle I describe below.
racist attitude tells us something important and negative about her—something for which she deserves blame. That is, we might think that Fiona is blameworthy for having a racist attitude. *Fellowship* involves an agent experiencing a certain attitude that is inappropriate and unjustified. Dan might plausibly be blameworthy for his unjustified resentment toward John as well as for the scowl that arises from this resentment.

*Birthday* and *Bad Joke* are different from *Sore Tooth* and *Fellowship* in that they involve some lack. Perhaps Angela lacks a certain attitude she should have—namely, some attitude regarding the importance of her friend, Dave, or the importance of keeping her promises. Additionally, Angela made a promise to Dave, so she ought to keep it. While Angela may protest that she really does care about Dave and values their friendship, the fact that she forgot to call him suggests otherwise. Although Angela did not choose or intend to forget to call Dave, it seems fitting for her to feel guilty for her omission and for Dave to feel some mild resentment toward her. Similarly, Ryland’s audience is justified in being indignant not only at her joke, which was voluntary, but also at her insensitivity, which was not. Ryland should be more aware of the impact her behavior has on others. Nevertheless, Ryland does not voluntarily ignore others’ welfare or opinions by choosing to ignore them; she simply fails to notice them.

None of these cases is extraordinary. In fact, they are commonplace. And yet our judgments that agents are responsible for their nonvoluntary faults seems to stand in tension with the commonly held principle that agents can only be responsible for what is voluntary, or as it is sometimes described, for what they choose, intend, or will. We are left with a puzzle about how to adjudicate our judgment that agents are responsible only for what is under their voluntary control and our judgment that agents are blameworthy for nonvoluntary faults. The question we must address, then, is whether these judgments that agents can be blameworthy for nonvoluntary faults (as well as our corresponding blaming practices) are justified. Can agents truly be responsible for the nonvoluntary, or for what they do not choose? If so, how?

### 1.2 Ways to Relieve the Tension

Here I explore three possible answers to this question. One answer takes a pessimistic view of our practices of holding others responsible:
(1) Our practices are not justified, and agents are never responsible for nonvoluntary faults.

Proponents of position (1) claim that agents are never responsible for nonvoluntary faults because these items are not things that we choose, intend, or will. Because choosing, intending, or willing to perform some action, \( A \), is necessary for an agent, \( S \), to be responsible for \( A \)-ing, \( S \) cannot be responsible for nonvoluntary faults. In other words, our practices are deeply flawed, as we regularly hold agents responsible for nonvoluntary faults without justification. Accepting (1) requires accepting that very many of our everyday responsibility practices and ascriptions are false. As this is a significant cost to bear, we should accept such a view only if we cannot justify these practices and ascriptions.

Perhaps because of this significant revisionary cost, others take a more subtle position that appeals to a distinction between direct and derived responsibility. When \( S \) is responsible for something, \( X \), in virtue of being responsible for something else that does not include \( X \), \( Y \), then \( S \) is derivatively responsible for \( X \). \( S \)’s responsibility for \( X \) is derived from \( S \)’s responsibility for \( Y \). For example, when Ted drunkenly hits and kills a pedestrian with his car, he is derivatively responsible for killing the pedestrian. Even though Ted does not choose to hit or kill the pedestrian, Ted chose to drink to excess when he was aware that he would have to get behind the wheel shortly. Because Ted freely, intentionally, and wittingly drank, Ted is responsible for drinking. In virtue of his prior choice to drink, he is derivatively responsible for hitting the pedestrian while drunk.

When \( S \) is responsible for \( X \) but not solely in virtue of being responsible for some other thing, \( Y \), \( S \) is directly responsible for \( X \). Sometimes this responsibility is called basic responsibility. Ted is directly responsible for his choice to drink, as Ted’s responsibility for his choice does not depend entirely on his responsibility for something else.

Appealing to this distinction between direct and derived responsibility, one might endorse the following position:

(2) Our practices are often justified, but agents can only be derivatively responsible for nonvoluntary faults, perhaps in virtue of their previous choices or willings, which indirectly led to the nonvoluntary fault in question.
Proponents of this view hold that we generally do not need to abandon our blaming practices. Some of these practices will be justified in virtue of the fact that agents chose or willed to act or omit in such a way that they would develop some nonvoluntary fault or other. As one might expect, however, the details here are vital. For example, Angela (from *Birthday*) may be only derivatively responsible for failing to remember Dave’s birthday and for failing to call Dave. In order to determine if Angela is responsible, we must know whether she chose to write down a reminder on her calendar. If she chose not to, she might be derivatively responsible for forgetting Dave’s birthday in virtue of choosing not to write a reminder. Her responsibility for failing to keep her promise would trace back to her choice not to write a reminder. But if she did write a reminder and still forgot, perhaps she is simply not responsible after all. She chose to do all she could reasonably be expected to do to remember her friend’s birthday.

The proponent of (2) claims that an agent is blameworthy for some nonvoluntary fault only if the agent’s responsibility can be traced back in some way, perhaps to a previous choice for which she is responsible. When there is no way to trace back responsibility, the proponent of (2) holds that the agent is not, after all, responsible for the nonvoluntary fault. As I will explore in chapter 4, this tracing strategy faces some difficulties, and it seems ill-equipped to explain why agents are responsible for certain nonvoluntary faults—despite the fact that we judge that such agents are responsible. Consequently, some of our judgments will likely have to be revised on this view. If there are a great number of cases for which the tracing strategy fails, however, (2) bears a significant revisionist cost as well. Though (2) will preserve more of our responsibility practices and ascriptions than (1), we might still wonder whether there is some other way to explain how our responsibility ascriptions and practices are justified that better matches how we hold agents responsible.

A third position denies that we need to trace back responsibility for nonvoluntary faults as in (2):

(3) Our practices are often justified, and agents can be directly responsible for some nonvoluntary faults (though not for *all* nonvoluntary faults).\(^4\)

---

\(^4\) A further position remains open, which is that an agent can be directly responsible for all of her nonvoluntary faults. But this view seems highly implausible. In fact, most of those who hold (3) are at pains to delineate precisely when (and why) agents are responsible for certain nonvoluntary faults so as to avoid this extreme position.
Those who favor position (3) claim that no previous choice or willing is necessary for an agent to be responsible for some particular nonvoluntary fault. This position includes several different theories, each of which may vary in explaining why and how agents can be directly responsible for nonvoluntary faults, as well as disagreeing about precisely which faults agents may be directly responsible for. For example, some proponents of (3) may claim that agents can be directly responsible for having or lacking particular attitudes, forgetting certain things, or unwittingly omitting to do something or other. These proponents may appeal to the fact that such nonvoluntary faults reflect attitudes that are sensitive to the agent’s judgments, and agents are directly responsible for their judgments. These judgment-sensitive attitudes reflect the agent’s commitments and thereby ground responsibility (Smith 2005). We might call such theorists liberal proponents of (3). Others may not be so permissive, however, holding that agents cannot be directly responsible for having or lacking attitudes, but can be directly responsible for omissions that are not intentional in virtue of certain capacities for voluntary control. These theorists are conservative proponents of (3).

I will call conservative proponents of (3), as well as those who endorse (1) and (2), volitionists. I will call liberal proponents of (3) nonvolitionists. Few writers endorse (1). The majority of volitionists seem to be those who favor a tracing strategy, as in (2). The chief dispute between these volitionists and nonvolitionists lies in where direct responsibility is to be found. Making progress toward settling this dispute requires that we address several questions: What, precisely, is the thesis that volitionists defend and nonvolitionists reject? What is voluntary control, and how might it be relevant to moral responsibility? Can responsibility for most of the nonvoluntary faults for which we hold agents responsible be accounted for derivatively, via tracing? What, if anything, justifies direct responsibility for nonvoluntary faults if they are not the products of our choices?

---

5 I follow convention in the literature here. Michael McKenna calls volitionists “voluntarists” (see McKenna 2008: 29 and 2012: 20). Nevertheless, “volitionist” seems to be the most common term (see Levy 2005, A. Smith 2005, and H. Smith 2011). Note also that volitionism regarding responsibility is a different thesis than volitionism regarding action theory. Though one may be both a volitionist regarding action as well as a volitionist regarding responsibility, one need not be. One might, for example, be a volitionist about responsibility while rejecting volitionism in action theory.

6 I break with convention here. Neil Levy 2005 and 2008, Holly Smith 2011, George Sher 2009, and others refer to certain theorists who reject volitionism as “attributionists.” As I explain below, however, I think this name is misleading, as it invites readers to understand responsibility in a particular way. Instead, I will refer to anyone who rejects volitionism as a nonvolitionist. As I understand these terms, one might be a nonvolitionist without being an attributionist, since it seems that attributionism is but one variety of nonvolitionism.
I turn to these questions in the chapters to come. First, however, I must address a more fundamental issue that some have seen to be at the heart of the dispute between volitionists and nonvolitionists: what it means to be morally responsible.

1.3 Kinds of Responsibility

Some have claimed that the debate between volitionists and nonvolitionists is not a substantial one, as volitionists appeal to one kind of responsibility and nonvolitionists to another. If this is right, it may be that volitionists and nonvolitionists are talking past one another.

Some have argued that there is a kind of responsibility called *attributability*. While it is not always clear what attributability amounts to, “as it is often characterized, an agent bears [attributability]-responsibility for something just in case that thing can be appropriately attributed to the agent as a basis for certain sorts of moral reaction to her” (Clarke, McKenna, and Smith 2015: 3). Many follow Gary Watson in his understanding of attributability. Watson holds that an agent is attributability-responsible for whatever expresses her practical identity or commitments. We judge agents as good or bad, virtuous or vicious, in response to these sorts of commitments that are revealed in an agent’s thoughts or actions (Watson 2004).

Watson contrasts attributability with *accountability*, which concerns demands and obligations such that when an agent has violated some such demand, she can rightly be subject to the reactive attitudes (e.g., resentment, indignation, or guilt) or possibly even sanctions. As Clarke, McKenna, and Smith point out, Watson’s understanding of accountability raises issues about fairness and our authority to impose sanctions that mere attributability apparently does not (2015: 4).

Some have rejected this distinction, arguing that responsibility should be understood as a unified concept—one of *answerability*. An agent is answerable for some item if we can sensibly demand from her the reasons that she took to justify that item. These reasons can be good or bad, and an agent may be subject to moral appraisal as a result. In fact, according to Smith, an

---

7 For example, see Levy and McKenna 2009: 118. It is unclear exactly what the metaphysical status of this distinction in responsibility is supposed to be. Some refer to different *kinds* of responsibility, while others appeal to *types, senses, or faces*. I don’t intend to take a stance on this issue here, but I will use “kind” for consistency.

8 Angela Smith 2012 and Pamela Hieronymi 2008 both argue for this kind of responsibility. Smith argues that answerability does all the work that other “kinds” of responsibility do, so answerability is the only kind of responsibility there is (2008, 2012).
agent may be appropriately subject to the morally reactive attitudes, including resentment, indignation, and guilt if the item in question is bad.

Notice that, according to some writers, what sets these kinds of responsibility apart from one another is, in part, the response that is fitting toward the agent. Imagine that Henry is running for senate, but his opponent, Miranda, has been doing well in the polls by slingling mud. When meeting with campaign staff, Henry might have the thought that he could simply have Miranda killed. He dismisses the thought as something that is off the table.\(^9\) One might think that Henry can be morally appraised as a bad person, or slightly dangerous. The fact that he had the thought to kill Miranda reveals his moral identity and his evaluative commitments. But, one might claim, it is not clear that any morally reactive attitude would be appropriate toward Henry. Such reactive attitudes are reserved for violations of expectations or obligations, but such obligations do not apply to thoughts in this way. For this reason, one might think that Henry’s thought about Miranda is attributable to him, but he is not accountable for it. We might morally appraise him as a bad person for having the thought but insist that he does not deserve any morally reactive attitude because he did not violate some obligation in having the thought.

This alleged difference in requirements and in fitting reactions to some agent that is reflected in different kinds of responsibility suggests that if there are kinds of responsibility, some kinds of responsibility are more robust than others. Kinds of responsibility that warrant the morally reactive attitudes, for instance, will be more full-blooded kinds of responsibility. If this is right, then the kind of responsibility at issue may be key in understanding the volitionist debate. It is tempting to think that volitionists and nonvolitionists are simply talking past each other, as nonvoluntary faults are clearly attributable to agents, for instance, but it is not clear that agents violate some expectation by having or lacking certain attitudes. One might plausibly think that, though nonvoluntary faults are attributable to agents, agents are not accountable for their nonvoluntary faults. The plausibility of nonvolitionism varies with the kind of responsibility one favors.

While the debate over whether responsibility is best understood as attributability, answerability, or accountability is an interesting one, it is not one that I will enter here.\(^{10}\) Nor do I think it necessary. Settling the question of whether we ought to understand responsibility as

\(^9\) This example is adapted from Williams 1985: 185.

\(^{10}\) For a glimpse of this debate, see Shoemaker 2011, Smith 2012, and Talbert 2012a.
attributability, answerability, accountability, or some combination of these seems to be putting the cart before the horse. For example, one might think that we should understand responsibility as accountability. Accountability is often understood as pertaining to an agent’s obligations. If obligation requires some sort of voluntariness or control that agents simply do not have over their nonvoluntary attitudes, then agents cannot be obligated to have these attitudes. If being accountable requires that an agent violate or fulfill an obligation, then one might think that agents can never be accountable for their nonvoluntary attitudes. Claiming that accountability is the true kind of responsibility at issue might seem to settle whether agents can be responsible for nonvoluntary faults, but only by way of stipulation about what the true nature of responsibility is rather than careful arguments. But things may not even be so simple as this. Nonvolitionists like Smith claim that agents are simply obligated to have or not to have certain attitudes (2008: 383, n. 19). If agents can be obligated to have certain attitudes, and accountability involves a violation of obligations, then it becomes unclear exactly what accountability requires over answerability such that we ought to make the distinction in the first place.

Instead of diving into these issues, I will focus on the merits of the volitionist and nonvolitionist positions first. After coming to some conclusions regarding that debate, we might turn to the question of whether responsibility is best understood as attributability, answerability, accountability, or some combination of these. Nevertheless, volitionists and nonvolitionists do need some shared notion of moral responsibility to avoid talking past each other. In fact, they already seem to have it. Both parties are interested in whether the morally reactive attitudes are justified or deserved with regard to some item. Neil Levy, who identifies as a volitionist, writes, “To say that an agent is morally responsible (for an act, omission or attitude) is to say that the Strawsonian reactive attitudes are justified in relation to her with regard to that act, omission or attitude” (2005: 1). Angela Smith, who rejects volitionism, writes something slightly different: “to say that someone is ‘morally responsible’ for some thing is to say that she is open, in principle, to moral appraisal—including moral praise and blame—on the basis of it (where nothing is implied about what that appraisal, if any, should be)” (2008: 370). But she is quick to clarify that she only avoids talk of the morally reactive attitudes to make room for morally neutral acts, which may not warrant any morally reactive attitude, and because “whether a Strawsonian reactive attitude is ‘justified’ in a particular context might involve considerations that have nothing to do with the agent’s responsibility and blameworthiness” (370, n. 9). Here,
Smith is thinking of considerations that are commonly cited in the ethics of blame, such as whether the blamer has the moral standing to blame. If the blamer lacks the moral standing to blame some agent, then having or expressing the morally reactive attitudes toward the agent may not be all-things-considered appropriate, even though the agent is blameworthy for some thing.\textsuperscript{11}

Pamela Hieronymi, too, claims that she is focused on responsibility of this kind:

To be \textit{morally responsible} for a thing (whether an attitude, an action, or a state of affairs) is to be open to moral assessment or judgment on account of that thing, to be open to moral praise or blame for it, to be open to what have come to be called the ‘reactive attitudes’ of resentment and indignation on account of it, and perhaps to be open to moral sanctions or punishments for it. (2008: 361)

Hieronymi goes further by including sanctions and punishments in her understanding of responsibility. I will focus only on the morally reactive attitudes.\textsuperscript{12} The kind of responsibility I take to be at issue in the debate between volitionists and nonvolitionists is whatever kind warrants the morally reactive attitudes. As I am focused on moral faults, I am particularly concerned with negative morally reactive attitudes.

I understand a negative morally reactive attitude to be an attitude some agent has in response to the belief that another agent has done wrong, has done something bad, or is morally bad in some way. The paradigm negative morally reactive attitudes are indignation, resentment, and guilt, though disapprobation and certain types of moral anger may also be negative morally reactive attitudes as I understand them.\textsuperscript{13} Given my characterization of the negative morally reactive attitudes, one can be indignant not only at another’s wrong action, but also at another’s

---

\textsuperscript{11} For more on the appropriateness of blame, see Smith 2007 and Coates and Tognazzini 2012 and 2013.

\textsuperscript{12} I take the morally reactive attitudes to be different from sanctions in that sanctions aim to cause suffering in another. The morally reactive attitudes can be used as sanctions when one expresses one’s reactive attitude with the intent to cause some harm. Since the morally reactive attitudes may be felt and expressed without any such intention, however, they are importantly different from sanctions. I return to whether punishment or sanctions might be justified for nonvoluntary faults, including objectionable attitudes, in section 7.3.4.

\textsuperscript{13} This is a departure from those who understand the morally reactive attitudes in a more restrictive way, such as Wallace 1994. Wallace restricts the morally reactive attitudes to only those attitudes that are in response to violations of obligations or expectations that one accepts. While there is much to be admired in Wallace’s restriction of the morally reactive attitudes in this way, I think Wallace restricts the reactive attitudes too narrowly. If volitionists reject that one can be obligated to simply have certain attitudes, they will reject that the reactive attitudes can apply to attitudes. This is why I construe the reactive attitudes more broadly. Even if agents can be obligated to have certain attitudes, this is a claim that requires defense, which I am not prepared to offer here.
objectionable attitudes. The important question, then, is whether one’s indignation in each case is justified.

When I say that an agent is responsible for something, I mean that the agent may be worthy of the morally reactive attitudes, so understood, on the basis of that thing. Of course, it may be that agents are not or cannot be worthy of the negative morally reactive attitudes for nonvoluntary faults, but this is a substantial question that must be addressed in its own right. The question to be answered, then, is whether some agent can be worthy of the negative morally reactive attitudes in a direct way for having certain nonvoluntary faults.

1.4 The Road Ahead

Addressing this general question will require tackling several others. In this introductory chapter, I have only gestured at the volitionist thesis. But very few have taken the time to carefully present and defend such a thesis. I begin to remedy this omission in chapter 2 by presenting what I take to be the best provisional formulation of the volitionist thesis. Rather than relying on voluntary control, as others have done, I suggest that we understand the volitionist thesis in terms of intentional actions. This affords us a clearer understanding of the debate between volitionists and nonvolitionists.

Because the volitionist thesis in chapter 2 is merely provisional, I explore a few different developments of the volitionist thesis in chapter 3—particularly whether volitionists can include omissions in the volitionist thesis in a way that is not ad hoc. Ultimately, I conclude that showing that agents are directly responsible for omissions is more problematic for most volitionists than has previously been appreciated.

If volitionists cannot accept that agents are directly responsible for certain nonvoluntary faults and omissions, might they account for these items derivatively via the tracing strategy? This is the topic of chapter 4. Ultimately, I argue that there are problems with the epistemic conditions on derivative responsibility that have been proposed, and some nonvoluntary faults—especially those of negligence as well as certain types of omissions—will prove quite difficult, if not impossible, for the tracing strategy. Accounting for responsibility for some nonvoluntary faults either directly or derivatively might be much more difficult for volitionists than they have

14 I will leave aside the issue of morally neutral actions here that concerned Smith 2008 above. I will also focus on whether an agent is blameworthy for some item, but not on whether the agent should be blamed for the item. The latter encompasses many considerations that are outside my focus, such as one’s standing or right to blame another.
previously appreciated. In light of this difficulty in matching our everyday responsibility ascriptions and practices, we have good reason to at least explore nonvolitionist accounts instead.

In chapter 5, I explain what I take to be among the leading nonvolitionist accounts currently available: Angela Smith’s rational relations view. I then defend this view against some common objections. In chapter 6, I address a recent criticism of views like Smith’s from Neil Levy. Levy argues that direct responsibility requires consciousness, yet the rational relations view seems to reject this condition on direct responsibility. I argue that Levy fails to establish consciousness as a necessary condition on direct responsibility, and then offer some revisions to the rational relations view that might help to mitigate the worries that Levy has raised. This modified version of the rational relations view is thus poised to be a viable alternative to volitionist views—especially in light of their apparent shortcomings. Finally, in chapter 7, I clarify the contours of this modified nonvolitionist view and defend it against some lingering objections. While I do not aim to show that no volitionist view can possibly serve as a viable account of our direct responsibility or that this modified nonvolitionist view is unassailable, I ultimately conclude that this nonvolitionist view can stand as a serious and legitimate contender to volitionist views in light of their shortcomings and problems. It is therefore worth developing further as we search for the best account of how and why agents are morally responsible for not only intentional actions, but also omissions and attitudes.
CHAPTER 2

THE VOLITIONIST THESIS AND VOLUNTARY CONTROL

2.1 Introduction

There is no single statement of the volitionist thesis. While volitionism seems to be a commonly held position, few theorists clearly identify themselves as volitionists. Those who do rarely offer an argument for some thesis that might be understood as paradigmatically volitionist. Nonvolitionists frequently discuss volitionism as a target, but they also fail to provide a precise statement of the view that they reject. Instead, they describe volitionism as the view that voluntary control or choice is required for responsibility, neglecting or setting aside explanations about how voluntary control and choice are related and why they are required for responsibility.

This lack of clarity regarding the volitionist thesis leaves us with only a general idea of the debate between volitionists and nonvolitionists. It is difficult to assess arguments for a thesis that is not clearly articulated, however. But formulating a precise volitionist thesis is no simple task. In this chapter, I offer such a formulation. In section 2.2, I examine typical characterizations of volitionism offered in the literature, settling on what I take to be the sort of volitionist thesis these writers have in mind. Because this thesis depends upon some conception of voluntary control, in section 2.3 I turn to the task of unpacking this notion, focusing on an account offered by Robert Adams. I argue that this account fails, so in section 2.4 I explore ways to remedy the volitionist thesis and the conception of voluntariness upon which it depends. Ultimately, however, I suggest that such remedies are unfruitful. We would do better to formulate the volitionist thesis in a way that does not depend upon some conception of voluntary control or voluntariness at all. In section 2.5 I offer just such a provisional volitionist thesis.

2.2 The Traditional Volitionist Thesis

To understand the volitionist thesis, let’s consider how volitionism is typically characterized:
C1: “[W]e are ethically accountable only for our voluntary actions and omissions.” (Adams 1985: 3)

C2: “Volitionists about moral responsibility…believe that a person is morally responsible only for those things that she has—directly or indirectly—chosen.” (A. Smith 2008: 369)

C3: “[C]hoice or voluntary control is a precondition of legitimate moral assessment.” (A. Smith 2005: 236)

C4: “[We can distinguish] two opposed accounts of moral responsibility, the accounts sometimes called attributionism, which stresses the extent to which an action is reflective of an agent’s real self, and volitionism, which stresses choice and control.” (Levy 2008: 214)

C1-C4 emphasize choice and voluntary control. Agents are only responsible for those items they choose, or those items with respect to which they have voluntary control. Given these characterizations of volitionism, one might naturally wonder what is meant by voluntary control, and how choice is related to voluntary control. Unfortunately, most theorists have either been entirely silent with respect to these questions or else have merely gestured at answers.

Michael McKenna has helpfully illustrated the lack of clarity that writers use when expressing the volitionist thesis. He notes that Angela Smith characterizes volitionism as restricting responsibility to “what an agent has voluntary control over,” or “whatever an agent has direct control over,” or “what an agent has chosen,” or “what an agent has voluntarily chosen,” or what an agent has deliberately chosen (McKenna 2008: 30). This lack of clarity is alarming, because these different characterizations of volitionism may capture different views, depending on one’s theory of action. McKenna writes:

[N]ot every choice needs to be understood as deliberate, especially if one takes ‘deliberate choice’ to mean a choice that is the upshot of deliberation. Also, what one has direct control over might be different from what one has voluntary control over. […]
Furthermore, the range of actions for which an agent is morally responsible could be construed to be much wider than the range of actions that involve choices. (2008: 30)

McKenna points out several distinct sorts of views, none of which is clearly marked as Smith’s target. Choice and deliberate choice are not necessarily coextensive, nor are direct control and voluntary control (though they may overlap). Without consistent language and an explanation of the role choice plays in voluntary control, it is unclear what it means to be a volitionist regarding responsibility. While this in itself is reason enough to push for a clear statement of the volitionist thesis, nonvolitionists have further reason to do so, since to understand the nonvolitionist position, one must understand what it is that nonvolitionists deny.

Clearly stating a volitionist thesis is complicated in part because volitionism is a family of views rather than one particular view. What appears to be a lack of clarity may simply be Smith’s attempt to capture as many of these volitionist views as possible: “What I have called the volitional view of responsibility is actually better understood as a cluster of distinct views which share a common assumption, namely, that choice, decision, or susceptibility to voluntary control is a necessary condition of responsibility (for attitudes as well as actions)” (Smith 2005: 238).

Smith identifies three types of volitionist views, all of which are united by this common assumption that choice is importantly related to responsibility. *Prior choice* views claim that we must “be able to trace the development of an attitude to a person’s own prior choices or decisions” in order for the agent to be responsible for the attitude (239). According to prior choice views, if some agent, Jeff, is responsible for his unsympathetic, cruel attitude toward others, this responsibility is derivative and in virtue of some previous choice that Jeff made. Perhaps in high school, Jeff was a kind and sympathetic person, but was often bullied. To defend himself against bullying, Jeff chose to act in unsympathetic and cruel ways towards others. Eventually his behaviors led him to have unsympathetic and cruel attitudes. While Jeff did not choose to have such attitudes, these attitudes can be traced back to his prior choices. This tracing technique is a common way to capture responsibility for attitudes.15

A second group of volitionist views is *endorsement* views. These views require that “the person herself endorse or identify with the attitude in question” (239). While Smith does not

---

15 I allude here to Manuel Vargas’s case of “Jeff the Jerk” in Vargas 2005. Fischer and Tognazzini 2009 respond to Vargas’s case in the same way a prior choice theorist would. I examine tracing in more detail in chapter 4.
elaborate on these sorts of views, presumably this endorsement must be a sort of conscious choice on the part of the agent to identify with certain attitudes. Without such a choice, it is unclear why the endorsement would count as volitional. Fiona in *Sore Tooth*, for instance, has not chosen to identify with her spontaneous racist attitude. Consequently, according to these endorsement views, Fiona is not responsible for her racist attitude.\(^\text{16}\)

The final category of volitionist views Smith calls *voluntary control* views, though I will refer to such views as *capacity* views.\(^\text{17}\) According to Smith, these views require “that a person have the ability to control or modify her attitudes through her future choices or decisions” (240). One may think that as Smith describes them, such views do not address retrospective responsibility, since the ability to make choices in the future seems to have no bearing on whether one is currently responsible for some action or attitude. But one might interpret these capacity views not as requiring the ability to control one’s actions or attitudes through *future* choices, but merely as requiring that one’s actions or attitudes be *susceptible* to one’s choices in some way.

The difference between prior choice views and capacity views is significant. The prior choice theorist holds that in order for *S* to be responsible for some nonvoluntary attitude, *A*, *S* must have chosen in such a way as to bring about *A*. The capacity theorist holds that in order for *S* to be responsible for some nonvoluntary attitude, *A*, it must be the case that *S* could have chosen so as to bring about or prevent *A*. (If *A* is an action, prior choice theorists require that *S* exercise control over *A* by choosing to *A*, while capacity views require merely that *A* was susceptible to *S*’s control such that *S* could have chosen to *A* or chosen so as to prevent *A*.) Capacity views then differ from prior choice views in that while prior choice views require that one have actually made some previous choice, capacity views require only that one had the *ability* (or capacity) to make some previous choice.

---

\(^{16}\) One writer Smith may have in mind as a proponent of endorsement views is Harry Frankfurt. But Frankfurt is not himself consistent regarding whether an agent must choose to identify with some attitude, so it is also unclear whether he holds that one must *choose* to identify with some attitude. In that case, it becomes much more speculative that Frankfurt is a volitionist of the endorsement stripe. In fact, Holly Smith 2011 has identified Frankfurt (1988: 8-94) as a nonvolitionist.

\(^{17}\) The reason for this is twofold. First, capacity views focus on an agent’s capacities, so the name is more fitting. Presumably all volitionist views are concerned with voluntary control in some way, so opting for “capacity views” over “voluntary control views” removes ambiguity. Second, it seems that some theorists who are currently defending these sorts of views refer to their views as capacity views. So this nomenclature better matches the practice in the current literature.
Prior choice views, endorsement views, and capacity views are united by a focus on the importance of choice or choosing to responsibility, but each family of views requires that choice be involved in a different way. One volitionist might think that endorsement views are too strong, as an agent need not choose to endorse certain attitudes in order to be responsible for those attitudes. Yet that same volitionist might think that some prior choice is required for responsibility, so that even if an agent does not choose to endorse some attitude presently, the agent is responsible for the attitude in virtue of prior choices she made. One can appreciate the difficulty, then, of offering a precise volitionist thesis that will set volitionists apart from nonvolitionists while leaving room to acknowledge the sometimes substantial disagreements between prior choice views, endorsement views, and capacity views.

While there may be no set of necessary and jointly sufficient conditions that can unite all volitionist families under one thesis, one might instead offer a weak, broad volitionist thesis of one necessary condition that sets volitionists apart from nonvolitionists. McKenna attempts to clarify Smith’s volitionist target by consolidating the key components of prior choice, endorsement, and capacity views. He offers the following as Smith’s target:

\[\text{Smith} \text{ has placed in her sights a paradigmatic formulation of the view that treats the voluntary as involving a species of direct control over action (presumably when the control is properly functioning), that takes the voluntary to be located in choice, and that conceives of choice as involving deliberate undertakings. (2008: 30)}\]

Voluntary control is but one type of control, and Smith sees volitionists as placing this type of control at the forefront of requirements for responsibility. But as McKenna importantly emphasizes, this voluntary control is a species of direct control over action in that this control is not mediated in any way. (I return to this point in the next section.)

Given these considerations, we might offer the following broad, unifying volitionist thesis:

\[\text{VTI: } S \text{ is directly responsible for some item, } x, \text{ only if } S \text{ exercised direct voluntary control over } x \text{ or } x \text{ was susceptible to } S \text{’s direct voluntary control.}\]

\[18\]

\[18\] I use the variable \(x\) here to stand for actions, omissions, attitudes, and consequences.
VT1 appropriately reflects the volitionist’s requirement that direct voluntary control is required for direct responsibility. Whether this control must be exercised or one must simply have some capacity for this control is left open in VT1. Although VT1 seems to capture what sets volitionists apart from nonvolitionists, it does not appreciate the important differences between prior choice and endorsement views, on the one hand, and capacity views on the other. The scope of responsibility on capacity views is wider than that of prior choice or endorsement views. To capture capacity views in a broad volitionist thesis, one will have to include items a prior choice theorist or endorsement theorist would likely ultimately reject. But restricting the thesis to satisfy prior choice and endorsement theorists will fail to capture items a capacity theorist includes within the scope of responsibility.

While a general volitionist thesis like VT1 that is concerned only with a necessary condition is useful, for the sake of clarity and prevision, we might instead separate VT1 into two separate theses, one of which captures more precisely prior choice and endorsement views, and another that captures capacity views:

*VT1a:* $S$ is directly responsible for $x$ only if $S$ exercised direct voluntary control over $x$.

*VT1b:* $S$ is directly responsible for $x$ only if $x$ was susceptible to $S$’s direct voluntary control.\(^{19}\)

Advocates of the prior choice view can endorse VT1a. According to McKenna’s understanding of Smith’s target, the voluntary is connected to choice. When $S$ chooses to perform some action, $S$ exercises direct voluntary control over that choice (and perhaps that action) and so can be responsible for it.\(^{20}\) Attitudes, however, are not things that we choose to have, and so not items over which we exercise direct voluntary control. Consequently, $S$ cannot be directly responsible

---

\(^{19}\) Note that VT1b includes all the items VT1a does, because if $S$ exercised direct voluntary control over $x$, then $x$ was susceptible to $S$’s direct voluntary control.

\(^{20}\) One might think that agents can only exercise direct voluntary control over their choices, not their actions. If one is not responsible for choosing, perhaps by exercising direct voluntary control over the choice, then it seems that one would not be responsible for the action that resulted from that choice. This suggests that direct voluntary control is limited to choices. While I find this argument persuasive, my goal is to be as accommodating as possible to volitionists here. As a result, I leave it open that agents might exercise direct voluntary control over both the choice to $A$ as well as over $A$-ing itself, depending on one’s account of direct voluntary control. For two accounts of direct control, see Mele n.d.
for some attitude on VT1a. Those who prefer endorsement views should also accept VT1a, provided they understand $S$ to be exercising direct voluntary control in endorsing some attitude.

Capacity views do not require that any actual choice be made in order for an agent to be responsible for some item, as long as the item was susceptible to the agent’s voluntary control. VT1b reflects this important difference between capacity views and prior choice and endorsement views. If $S$ could have chosen to $x$ and could have brought $x$ about by so choosing, then $S$ can still be morally responsible for $x$-ing or not $x$-ing—even if $S$ never made any actual choice to $x$ or not $x$. The ability to choose rather than the actual choosing is what is important in VT1b.

Whether one favors VT1 or the pair of VT1a and b rests upon one’s goal in understanding volitionism. A broad, weak thesis like VT1 helpfully separates volitionists from nonvolitionists, but does a poor job distinguishing important differences between volitionists. VT1a and b make the differences between volitionist views clearer, but offer less in the way of separating volitionists and nonvolitionists generally. While both of these goals are important, I will frequently discuss VT1a or b because they allow for precision that will be useful in appreciating differences in how different volitionists explain some cases.

While VT1a and b seem to be the sorts of volitionist theses that Smith and other nonvolitionists target, VT1a draws more attention than VT1b in the literature. In fact, Smith and other nonvolitionists say little to attack VT1b. Consequently, I will focus my attention largely on VT1a in this chapter, though I will discuss the importance of capacity views in later chapters. Regardless of which thesis one favors, VT1a and b remain obscure without some account of voluntary control. I turn my attention toward uncovering some such account in the next section.

### 2.3 An Account of Voluntary Control

McKenna suggests that Smith understands the voluntary as stemming from choice. Robert Adams, one of the first theorists to seriously question volitionism, also understands voluntary control at least partially in terms of choice. He offers the following account of direct voluntary control:

$VCd$: “To say that something is (directly) within my voluntary control is to say that I would do it (right away) if and only if I (fully) tried or chose or meant to do so, and hence
that if I did it I would do it *because* I tried or chose or meant to do it, and in that sense voluntarily” (1985: 8).

As Adams writes, “the key to this conception [of voluntary control] is the idea of a trying, choosing, or meaning of which what is voluntarily done is (or would be) the object” (8). He focuses on choosing at least in part because “in a voluntary action the action itself is an object aimed at by the agent in a way that is aptly described as ‘trying, choosing, or meaning’” (8).

VCd seems to give the right result for many cases of intentional actions. Suppose Sam raises her arm. According to VCd, to say that Sam’s arm raising is within her direct voluntary control is to say that she would raise her arm just in case she fully tried or chose or intended to raise her arm, and she would raise her arm because of this trying, choosing, or intending. If some manipulation of Sam’s brain (perhaps during brain surgery) causes her arm to rise, we would not count this as voluntary. Appropriately, VCd does not classify such manipulated arm raising as an exercise of voluntary control, as it would not occur as a result of Sam’s trying, choosing, or intending to raise her arm.

While VCd nicely captures how intentional actions are within one’s direct voluntary control, it also illustrates how one’s attitudes and emotions are not. One cannot, for instance, choose or intend or try to become angry and thereby make herself angry right away. Of course, there is a wrinkle here. One can engage in certain behavior in order to thereby make oneself angry. Suppose that Val decides to dwell on her stressful workplace where she is overworked, underpaid, and underappreciated, knowing that so dwelling will make her angry. In a sense, Val’s becoming angry is voluntary, but it is not direct. She does not choose or intend or try to become angry and thereby make herself angry right away. Val’s anger is only within her indirect voluntary control.

---

21 I understand “meaning” and “intending” to be equivalent, though Adams himself is unclear on this point. For consistency I will use “intending”.

22 There do seem to be counterexamples to VCd, however. Some cases will be strange or rare, such as cases of causal deviance. Perhaps a case in which an individual chooses to act and then immediately dies could also present a problem for VCd. Less rare are cases in which one chooses to *A* and *As* because one so chooses, but would have *A*-ed even if one hadn’t chosen, tried, or intended to *A*. For instance, Randy, sitting at his desk, might choose to remain sitting at his desk and remain there as a result of his choice. But if Randy hadn’t chosen to remain at his desk, perhaps he would have continued deliberating and would have remained sitting at his desk without trying or intending to do so as he deliberated. Such cases create problems for the necessary condition of the biconditional in VCd. (I owe this sort of counterexample to Randolph Clarke.)
To contrast with direct voluntary control, Adams offers a conception of indirect voluntary control as well:

\[ VC_i: \text{“Doing } x \text{ is not directly but only indirectly within my voluntary control if I could try successfully to do } x \text{ right away, but only by trying successfully to do something else as a means to doing } x \” \] (8, n. 5).

As we saw in chapter 1, volitionists hold that agents are only derivatively responsible, if responsible at all, for their attitudes. Thus the focus for nonvolitionists is on direct voluntary control as spelled out in VCd rather than the extended sense of “voluntary” involved in VCi.\(^{23,24}\) Because volitionists are focused on direct voluntary control, and because attitudes and emotions are not under one’s direct voluntary control according to VCd, attitudes and emotions are not something for which agents are directly responsible according to volitionists. VCd gives us the right result.

Note also that what agents forget or neglect is also not under their direct voluntary control. At best, one can bring about that one forget something by, say, choosing to not leave reminders or choosing to distract oneself with other activities. But one cannot simply choose to forget or neglect something and thereby forget or neglect it.

Despite these results, VCd is problematic as an account of direct voluntary control. Volitionists do not only want to capture intentional overt actions as voluntary. They also want to capture a variety of mental actions, including choosing and willing. Choosing and willing seem

\[ ^{23} \text{Pamela Hieronymi makes a similar point when discussing whether believing is voluntary. She notes that if one includes in the class of ‘the voluntary’ “not only voluntary activities but also those things that are the (possible) result of such activity” (2008: 358, n. 1), believing can be the result of voluntary activity. But the debate between volitionists and nonvolitionists lies in where direct responsibility is to be found. Hieronymi elaborates regarding responsibility for belief:} \]

\[ \text{[T]he special problem with responsibility for belief arises because it seems that we are somehow more directly responsible for our beliefs—responsible for them in a way beyond the way in which we are responsible for the management of our allergies or the state of our apartment—even though they are not themselves voluntary activities (not, as it is sometimes put, ‘directly’ voluntary). (2008: 358, n. 1)} \]

\[ \text{Just as one can voluntarily take steps to make oneself angry, one can voluntarily take steps to make oneself believe } p. \text{ But Hieronymi claims that we often think that an individual is more directly responsible for that belief, not for the steps that he took to form the belief.} \]

\[ ^{24} \text{One might worry that VCi is too weak. Recall from chapter 1 the case of Ted, the drunk driver who hits a pedestrian. It seems that Ted exercised his direct voluntary control in drinking alcohol. But Ted did not drink and get behind the wheel as a means to successfully hit a pedestrian. Nevertheless, it seems that Ted’s hitting the pedestrian was within his indirect voluntary control. I take this to be an indication that VCi needs refinement. As I will ultimately reject VCd, I do not attempt any refinement of VCi here, but merely note this problem.} \]
to be within one’s direct voluntary control. If we are to understand direct voluntary control as in VCd, however, we will be forced to say that S’s choosing to x is within her direct voluntary control just in case S would choose to x iff S tried or chose or intended to choose to x. This is problematic for two reasons. First, it is often false that in choosing to x, S chooses to x because S chose to choose to x. This would lead to an unwelcome regress. More importantly, however, we do not think that choosing is under one’s voluntary control just in case one chooses to choose.

While Adams does not acknowledge these problems, he does seem aware that VCd will have problems with choosing, trying, and intending. He writes:

Trying, choosing, or meaning can itself be regarded as voluntary, as an operation of the will, so to speak, that is characteristic of voluntary action. It may perhaps also be said, in an extended sense, to be within our voluntary control, though it is itself the controlling, rather than an object of control. (9)

This extension of VCd to cover trying, choosing, and intending will not do, however. In the first place, the extension seems ad hoc. Adams began with things that one does by choosing and then attempted to extend the concept of direct voluntary control backwards to cover choosing as well, despite choosing not satisfying the original parameters of his account. Secondly, it is not clear that trying, choosing, and intending should all be regarded as voluntary. As Holly Smith writes, “[m]any people believe that desires and intentions are not typically themselves voluntary, but ultimately arise from our biology, our environments, and our prior experiences” (2011: 123).

Additionally, one can non-actively come to intend to A. When Al unlocks his office door in the morning, he intends to unlock his door. But he does not actively acquire this intention. He is not uncertain about whether to unlock the door, so there is no need for him to actively form the

---

25 It seems that any conception of voluntary control that begins with such items will fail to capture vital items as within our voluntary control. Pamela Hieronymi attempts to cash out the voluntary by relying chiefly on intention: [A]n activity is voluntary just in case you decide to do it for reasons you take to settle the question of whether to do it, therein intend to do it, and providing all goes well, do it by executing that intention. That is, an activity is voluntary just in case one engages in the activity by forming and executing an intention to do so, where one forms an intention to engage in the activity in settling the question of whether to engage in it, where that question could be settled by any set of considerations that one takes both to count in favor of the activity and to be sufficiently strong (or, indeed, for no particular reason) (2008: 366). Hieronymi’s conception of the voluntary seems to rule out choice as well, as one often does not decide to choose to A. While this conception of the voluntary may suffice in some contexts, it fails as a useful conception for volitionists or nonvolitionists, who are interested in construing direct responsibility in terms of the voluntary.

26 In a similar vein, Pamela Hieronymi 2008 argues that intending is not voluntary.
intention to unlock the door (Mele 2003: 200-201). In this case, in coming to intend to unlock the door, Al has not voluntarily done anything. If this is right, then not all of these items should be understood as being within our voluntary control. Thus Adams cannot simply extend VCd to include choosing, intending, or trying. Nonactively acquired intentions are not voluntary, though the actions that issue from these intentions may be voluntary.

To summarize, VCd cannot be an accurate conception of direct voluntary control. Choosing, a paradigm action for which we are directly responsible and which is an exercise of our voluntary control, is not captured on VCd. But any extension of VCd to capture choosing, intending, or trying is \textit{ad hoc} at best, and such an extension will require further refinements to rule out nonactively acquired intentions that should not count as voluntary.

\section*{2.4 Choosings and Acts of Will}

According to Adams, “trying, choosing, or meaning can itself be regarded as voluntary” (1985: 9). But as I just argued above, this does not seem to be quite right. Because voluntary control seems to rely on a notion of the voluntary, we might first simply focus on what it means to be voluntary. Since we are focused on voluntariness in a direct sense, we should focus on what it means to be directly voluntary. Getting clearer on this notion should help us get clearer about what it means for something to be within our direct voluntary control.

If C1–C4 are any indication, choosing appears to lie at the heart of voluntary action. Given the lesson of VCd’s failure, our conception of the directly voluntary should not begin with the things one does by choosing, but rather with the choosing itself, lest we be forced to usher choosing through the back door once again. But not only are our choices voluntary, as Adams notes. The things we do by choosing are also voluntary. This gives us the following:

\begin{equation*}
VYd: x \text{ is directly voluntary for } S \text{ if and only if } x \text{ is some choice of } S \text{'s or something that } S \text{ does by choosing to } x.
\end{equation*}

\textit{VYd} offers an account of the voluntary in terms of choice, and extends the account to the things one does by choosing rather than beginning with the things one does by choosing and extending the voluntary to choice. By focusing on \( S \)'s choice as the key rather than the things that \( S \) does by choosing, \( VYd \) is an improvement over VCd.
Consider the following example from Donald Davidson: “I flip the switch, turn on the light, and illuminate the room. Unbeknownst to me I also alert a prowler to the fact that I am home” (1980: 4). On a coarse-grained view of actions, Davidson performs one action under four different descriptions. On a fine-grained view of actions, Davidson performs at least four different actions (see Goldman 1970). Suppose that Davidson chooses to flip the switch. Does he also choose to alert the prowler to the fact that he is home?

On a fine-grained view of actions, it seems easy to understand why Davidson did not choose to alert the prowler. Flipping the switch and alerting the prowler are two different actions, because they have different act-properties. The property of flipping the switch is distinct from the property of alerting the prowler, as one can have the former without the latter (and vice versa). Even on a coarse-grained view of actions, however, Davidson does not choose to alert the prowler. This is so even though Davidson performs only one action. On a coarse-grained view, actions simpliciter are not performed intentionally or for a reason. Only actions under some description are so performed. Davidson chooses to perform some action under the description “flips the switch” but not to perform some action under the description “alerts the prowler”. This is plausibly because choosing has built into it some kind of awareness such that one cannot choose to x unless one is aware of what x is.

I do not intend to take a stance on whether actions should be understood in a coarse-grained way or a fine-grained way here. Whether one takes a coarse-grained view of action or a fine-grained view of action, one can sensibly say that Davidson does not choose to alert the prowler. He voluntarily flips the switch, and he likely voluntarily turns on the light and illuminates the room. But he does not voluntarily alert the prowler. VYd can account for this subtlety.

Despite accommodating this subtlety, VYd is too narrow. Choices and the things one does by choosing are not the only things that are directly voluntary. We can try to do x, will to do x, and intend to do x—all without choosing to do x. In fact, most of our everyday behavior is unthinking and habitual, involving no choice at all, as in the case of Al habitually unlocking his door without choosing or deciding to unlock it. He nonactively acquires the intention to unlock his door and thereby unlocks his door. Only if Al became uncertain about whether to unlock the door (perhaps because he heard fighting inside) would Al then choose to unlock the door (Mele
Despite the fact that Al does not choose to unlock his door, Al voluntarily unlocks his door. Thus VYd is too narrow.

It is unclear, however, how to expand VYd. Perhaps we should focus not on choice, but on all acts of will. Choosing is but one paradigm instance of an act of will, but agents act voluntarily via other sorts of acts of will. This suggests the following:

\[ VYd^*: \text{x is directly voluntary for } S \text{ if and only if x is some act of will of S’s or something that } S \text{ does by some act of will.} \]

There are two problems facing VYd*. First, depending on how broadly we should understand the “by” in “something that S does by some act of will,” VYd* might count as directly voluntary things that one did not intend to do by some act of will. For instance, Ana might choose to go for a walk and go for a walk because of that choice. Walking is something that Ana does by some act of will—namely, her choosing. But in walking, Ana may also step on ants and wear down the soles of her shoes. These are things that Ana does by her act of will of choosing to go for a walk, in a broad sense. Yet Ana does not intentionally step on ants or wear down the soles of her shoes, and it seems odd to claim that such items are voluntary for Ana. Proponents of VYd* might try to restrict “by” in some way to rule out such cases. Perhaps revising the phrase to “something that S intentionally does by some act of will” would solve this problem.

The second problem with VYd* is that although choosing clearly counts as an act of will, it is far less obvious what else might be included. Some candidates might be wanting, trying, or intending. Of course, if acts of will are to be taken at face value, they will presumably be actions. But wanting is not an action that one performs; rather, one is in the state of wanting something. Perhaps this is sufficient to rule out wanting from being directly voluntary, though one might take steps to foster a desire such that wanting (or perhaps coming to desire) could be considered indirectly voluntary. Nevertheless, I take wanting to be ruled out as an act of will.

Trying is a more difficult case. While trying generally has a connotation of making some special effort or struggling against some difficulty, I understand trying here as many action theorists do. On this understanding, trying is involved in all intentional action. With Adams and

---

27 I understand acts of will and willings to be equivalent here, but opt for the term ‘acts of will’ for stylistic reasons.
Mele, I hold that “trying to A is an event or process that has A-ing as a goal and is initiated and (normally) sustained by a pertinent intention” (1992: 326). Successful tryings to A are A-ings.

Given that trying to A is initiated by some intention, it seems that trying should be understood as an act of will as well. But Randolph Clarke and Thomas Reed argue that only beginning portions of such attempts should be understood as acts of will. During the early portions of the trying process, one’s having an intention activates certain motor neurons, and this may rightly be said to be an act of will. But the entire trying is not an act of will. Clarke and Reed offer an example of a golfer trying to sink a putt. While the golfer’s hand movements might rightly be said to be part of the attempt to sink the putt, they argue, her movements do not count as part of her act of will. Hand movements, unlike a portion of the willing, may be slow. As the predicate appropriately applies to hand movements but not to portions of an act of will, the hand movements cannot be part of the act of will (Clarke and Reed 2015: 10-11).

Of course, some disagree. While it seems fairly clear that at least some portions of tryings or attempts should be understood as acts of will, it is far less clear what portion of the trying should be understood as an act of will. What gets included as an act of will may impact how we understand what S brings about by some act of will. Clarifying this dispute becomes important for those who wish to endorse VYd*.

Adams claims that intending is voluntary, so perhaps he would claim that intending should be included as an act of will. As I noted above, Pamela Hieronymi argues that intending is no more voluntary than believing (2008: 368). But even if Hieronymi is correct, one might think that forming an intention is an act of will. After all, sometimes one comes to have an intention to x by deciding to x, and deciding is an intentional action. Yet many intentions are formed nonactively, such as Al’s intention to unlock his office door each morning. It would seem, then, that not all instances of intention formation are instances of acts of will. Perhaps only those that are formed via decisions or choices will count, but we have already included choices as paradigm cases of acts of will.

There may be further candidates as acts of will. One might insist that volitions are acts of will. Or one might insist that one may only try when one may fail, and certain sorts of mental

---

28 For instance, O’Shaughnessy 2008 holds that the golfer’s hand movements should be understood as part of her act of will.
30 For instance, see Zimmerman 1988 and Ginet 1990.
actions like imaginings do not include this opportunity for failure.31 If so, then the way that I have described tryings above is incorrect, and there is some further sort of act of will. These issues are complex, and they threaten to splinter volitionists even further given their favored understanding of what count as acts of will.

VCd, VYd, and VYd* all face serious challenges. Rather than attempt to give some account of voluntary control or the voluntary, we might instead abandon VT1 and VT1a altogether. Instead, we can formulate the volitionist thesis without relying on such notions as voluntary control and the voluntary.

2.5 Starting Over: A New Statement of the Volitionist Thesis

The failures of VCd, VYd, and VYd* were the result of being too broad or too narrow. VYd* includes choice, but it is not clear what else it ought to include. As McKenna observes, Smith and other theorists understand “the voluntary to be located in choice,” and they conceive of choice “as involving deliberate undertakings” (2008: 30). While choice involves a deliberate undertaking, we have seen that this is too narrow to capture everything a volitionist plausibly wants to capture. One reason for this may be that volitionists and nonvolitionists alike are using “choice” and “choosing” in a much broader sense than is common in action theory. Instead, they seem concerned with deliberate undertakings in general. After all, acts of will such as choosings are intentional actions. If this is right, rather than focus on choice and voluntariness, we might get a better grasp on the debate by focusing instead on intentional actions.

Consider the following thesis:

VT2: S is directly responsible only for S’s intentional actions.32

VT2 is far cleaner and clearer than its predecessors. There is no concern about using “choice” in an ambiguous way. Unlike VT1a and b, VT2 requires no account of voluntary control. And

31 This is what O'Shaughnessy 2009 argues.
32 Notice that VT2 offers only a necessary condition, as VT1 and VT1a and b did. Perhaps there are further conditions a volitionist wants to include to capture sufficiency. For instance, one might think that S is responsible only when S’s intentional actions are free. VT2 is compatible with S being directly responsible only for some subset of S’s intentional actions, though which other conditions a volitionist may desire to add will depend on what distinct volitionist view one endorses.
unlike VCd, VYd, and VYd*, it is clear that only intentional actions are included under the scope of direct responsibility.

To better understand VT2, one might sensibly ask for an account of intentional action, just as we sought an account of voluntary control to better understand VT1a and b. I adopt here the account of intentional action offered by Alfred Mele and Paul Moser 1994. Mele and Moser give necessary and sufficient conditions, but two of these conditions are designed to handle cases of lucky action. I am not concerned with these sorts of cases here, and so focus only on two necessary conditions that are distinctive of intentional action: “Necessarily, an agent, S, intentionally performs an action, A, at a time, t, [only if] (i) at t, S A-s and her A-ing is an action, [and] (ii) at t, S suitably follows—hence, is suitably guided by—an intention-embedded plan, P, of hers in A-ing” (Mele and Moser 1994: 63).

The first condition is rather trivial, but the second warrants further explanation. In performing an intentional action, one has an aim to perform an action of some type. Mele and Moser note that “Our performing an action of type A intentionally requires our having some representation of that action type, or at least some representation that can suitably guide our performing an action of that type” (44). This representation, as Mele and Moser understand it, is an action plan, which may be relatively simple (a representation of one’s A-ing) or more complex (a representation of a plan including the means by which one will achieve some goal) (42). This representation must guide the action in order for the action to be intentional.

Like VT1a and b, VT2 entails that agents can be directly responsible for many items that volitionists wish to capture under the scope of direct responsibility. Most notably, mental actions like choices are intentional actions. While intending is not an intentional action, actively forming an intention will be included as an item for which agents may be directly responsible.

---

33 These are conditions (iii) and (iv) of Mele and Moser 1994: 63.
34 Some may think that in settling on VT2 as the volitionist thesis, I have simply returned to a notion of the voluntary or voluntary control. Hieronymi writes that intentional actions “are, in a certain sense, ‘voluntary’: a person decides to perform an intentional action by answering for herself (however implicitly) the question of whether so to act—a question that includes in it a representation of that action, under the relevant description” (2004: 118). There is no conflict here between this thought and the analysis of intentional action offered by Mele and Moser 1994. I take it that when volitionists focus on voluntary control, they are concerned with something like intentional action. Often, talk of the voluntary can be rephrased in terms of intentional actions. For instance, Holly Smith writes, “the voluntary is that which proceeds from, and carries out, the will” (2011: 125). One might interpret Smith as saying that the voluntary is that which proceeds from some intention and carries out some action plan sustained or guided by the intention. If so, so much the better for VT2, as it does not stray too far from VT1 and yet it is far clearer, as it makes no appeal to some mysterious “will”. Finally, it is important to note that although intentional action and voluntariness may be closely related, one need not appeal to the voluntary or voluntary control in an analysis of intentional action.
Additionally, a variety of intentional actions, such as raising an arm, singing a song, driving a car, or killing someone might also be included within the scope of direct responsibility.\textsuperscript{35}

VT2 excludes from direct responsibility many items that volitionists want to exclude. Attitudes, for example, are not actions, so they are not intentional actions. Consequently, they are not items for which one may be directly responsible. Of course, this is not to say that one cannot be directly responsible for intentional actions one performs to bring about some attitude. Additionally, forgetting is not an intentional action, and so is not something for which agents are directly responsible according to VT2. Again, one might perform intentional actions to bring it about that one forgets something. Such intentional actions may fall under the scope of direct responsibility, according to VT2. But forgetting itself will not count as something for which agents can be directly responsible.

While VT2 is cleaner and avoids the problems of VT1, one may worry that it is too narrow. VT2 rules out many omissions from the scope of direct responsibility. While philosophers do not agree on what omissions are, Clarke 2012 argues convincingly that many omissions are simply absences of action. Some omissions may be negative actions (e.g., doing one thing to prevent oneself from doing another), while others may be positive actions linguistically described in a negative way (e.g., holding still). But not all omissions can be accounted for in this way. Some simply are absences of action. If this is correct, then VT2 rules out direct responsibility for a great deal of omissions, as these absences are not actions, let alone intentional actions.

At first, the fact that VT2 excludes certain omissions from the scope of direct responsibility might seem sufficient reason to reject it as a useful volitionist thesis. Notice, however, that even though certain omissions are not captured in VT2, this does not imply that agents cannot be responsible for those omissions. VT2 is a thesis about \textit{direct} responsibility. Perhaps agents can only be \textit{derivatively} responsible for these omissions. With some additional requirements regarding derivative responsibility, these absences of action might be accounted for.

\textsuperscript{35} Importantly, however, they need not be since VT2 is only a necessary condition on direct responsibility. Some volitionists may hold that agents are directly responsible only for basic actions, in which case some of these intentional actions might not be included. Nothing in VT2 forces a volitionist to include these items, however; VT2 simply restricts the class of items for which agents can be directly responsible. This class might sensibly be restricted even further depending on what other commitments one holds. I discuss this point in more detail below when considering VTZ.
Nevertheless, one might prefer to account for such omissions directly. There seem to be two salient ways to do this:

\[VT3: \text{S is directly responsible only for intentional actions and omissions of intentional actions.}\]

\[VT4: \text{S is directly responsible only for intentional actions and intentional omissions.}\]

While both VT3 and VT4 include omissions, they differ in which omissions they include. VT3 is wider than VT4, as it does not require that some omission be intentional. Randy might promise his wife that he will buy milk on the way home from work, yet nevertheless forget to stop on the way home. Randy’s omission seems to be an absence of action, but more specifically, it is an absence of an intentional action. Yet Randy’s omission is not an intentional omission; he did not intentionally omit to get the milk. He simply forgot.\(^{36}\)

Because VT3 is broader than VT4, it corresponds more closely with capacity views. Randy had the capacity to pick up the milk on the way home, and therefore he may be directly responsible for failing to do so. VT4 corresponds with prior choice and endorsement views, as these views require one to exercise some kind of control, perhaps by choosing to omit to pick up the milk. VT3 is a broader thesis. It is compatible with all volitionist views, and it does not seem to be a thesis that nonvolitionists would accept. As a result, it can serve as a way to separate volitionists from nonvolitionists (though see note 36). But VT4 is more precise in capturing what is distinct about prior choice and endorsement views.

Because VT3 is merely a necessary condition, proponents of even the most restrictive volitionist views can accept it. Nevertheless, these more restrictive views will look quite different from volitionist views that are more inclusive. For example, Michael Zimmerman explicitly endorses a thesis like the following:

\[^{36}\text{This case comes from Clarke 2014: 164 ff. Notice just how broad VT3 is. It leaves open the possibility that one is directly responsible for unintentional, nonvoluntary omissions. Yet responsibility for these sorts of unintentional omissions is one of the chief motivations nonvolitionists offer for their view (see cases in Smith 2005). For example, recall Birthday. Prior choice theorists either deny that Angela is responsible for not calling her friend, or they claim that she is derivatively responsible in virtue of some previous intentional omission or action. (Levy 2005 replies in this way.) But VT3 might allow that Angela is straightforwardly directly responsible for her unintentional omission, as it is an omission of an intentional action. Many volitionists might balk at such claims.}\]
VTZ: S is directly responsible only for S’s volitions.

Zimmerman holds that “an agent is directly free only with respect to his volitions, so he is directly culpable only for his volitions” (1988: 40). He understands a volition to be “a decision or choice (not necessarily based on deliberation but still practical in nature and made in light of certain background beliefs) that some event occur, a decision which is accompanied by an intention that it (the decision) be causally efficacious with respect to the event in question” (19). For Zimmerman, volitions are types of decisions, though not all decisions are volitions.

Zimmerman’s understanding of volitions severely limits the scope of direct responsibility. Many theorists will hold that agents can be directly responsible not only for their choices, but also for the actions or omissions they perform by choosing. Perhaps agents can even be responsible for certain routine, “thoughtless” actions, such as Al’s opening his office door. It is difficult to see how these actions are included in VTZ. If the goal is to find a volitionist thesis that separates volitionists from nonvolitionists, however, VT3 seems a plausible candidate, as it includes within it theses like VTZ. Nevertheless, a thesis like VT3 obscures important differences between volitionists, and some volitionists may think it is too broad to be useful, as it does not express what is distinct about their views. I aim to offer a thesis that separates volitionists from nonvolitionists, but I also prefer to offer more precision if possible. This is why offering VT4 as an alternative to VT3 is useful, even though VT4 excludes capacity views.

---

37 Zimmerman claims that VTZ does not rule out responsibility for omissions, as an intentional omission “cannot be accomplished…without intentionally doing something in the place of that which one omits to do. Thus all intentional omission indirectly involves volition” (1988: 23). Zimmerman appeals to the following: “If S intentionally omits to bring about e, then there is some event f such that S wills f for the purpose of not-e, where ‘not-e’ signifies a ‘negative event,’ viz., the negation of e” (23). While I cannot argue the point here, Zimmerman’s understanding of omissions seems incomplete, as some intentional omissions will not be negative events at all, but simply absences of action. For more on this, see Clarke 2010a and 2012.

Additionally, Zimmerman allows that an agent is “indirectly free with respect to the consequences of those volitions with respect to which he is directly free” (1988: 7). If agents can be derivatively responsible for that with respect to which they are indirectly free, then agents can be derivatively responsible for the products and consequences of their volitions. Zimmerman allows for one to be responsible for what one decides and intends to bring about, but traces that responsibility to volition.

38 Zimmerman claims that “all action, even unthinking and habitual action…involves a volition qua decision, even though the agent is oblivious (or, at least, not fully, occurrently mindful) of the volition in question” (1988: 21). While I will not argue the point here, I find this claim highly implausible. It is unlikely that Al has some volition in Zimmerman’s sense when Al habitually unlocks his door, as Al does not choose to unlock the door. Like Mele, however, I understand choice (or decision) to involve the settling of some uncertainty. Perhaps Zimmerman is using “choice” in a different way here.
Whether volitionists can accept VT3 or VT4 over VT2 is a substantive question, and one that I will turn to in chapter 3. Before addressing this question, however, it is worth looking briefly at another writer who has also attempted to state the volitionist thesis without appealing to voluntary control or the voluntary. Instead, Michael McKenna focuses on the control one is able to exercise by acting freely:

\textit{VTM}: “A person is morally responsible only for things that are within the control she is able to exercise by acting freely” (2012: 187).

While I think that McKenna is right to abandon the focus on voluntary control and the voluntary, I do not think VTM will serve as a clear volitionist thesis. Notice that VTM does not distinguish between direct and derived responsibility. This obscures the debate between volitionists and nonvolitionists. Consider those who hold prior choice views. They agree that one’s attitudes may be within the control one is able to exercise by acting freely. But, of course, nonvolitionists accept that one is responsible for these items as well, though not \textit{in virtue} of the fact that those items were within the control one was able to exercise by acting freely.

Nevertheless, perhaps VTM might be modified to account for the distinction between direct and derivative responsibility. If so, VTM might be comparable to VT3 provided that McKenna is correct about how VTM handles omissions. McKenna claims that “the same control one exercises by acting freely is the control implicated in omissions. In the case of the relevant class of omissions, one refrains from exercising that control” (2012: 188, n. 11). One might omit from exercising this same type of control when one forgets to pick up the milk on the way home or call a friend on his birthday, as promised. I prefer VT2, VT3, or VT4 to VTM because it makes no appeal to control. It is reasonable to ask a proponent of VTM just what kind of control agents are able to exercise by acting freely. If this control is understood as voluntary control, familiar problems may arise once again.
CHAPTER 3

THE VOLITIONIST THESIS AND OMISSIONS

3.1 Introduction

In chapter 2, I argued that the volitionist thesis should not be spelled out in terms of voluntary control. Instead, I offered what I take to be a better understanding of the volitionist thesis:

\[ VT2: S \text{ is directly responsible only for } S\text{'s intentional actions.} \]

VT2 rules out omissions that are absences of action from the scope of direct responsibility. If we are not directly responsible for these omissions, some volitionists might suggest that we are derivatively responsible for them, perhaps in virtue of some previous intentional action such as a choice to omit, refrain, or forbear. Many of our omissions, however, like many of our actions, are not preceded by some choice. We are not uncertain about what to do, but simply act or omit straightaway. For this reason, the volitionist may want to include omissions within the scope of direct responsibility.

I suggested two ways of doing this:

\[ VT3: S \text{ is directly responsible only for intentional actions and omissions of intentional actions.} \]

\[ VT4: S \text{ is directly responsible only for intentional actions and intentional omissions.} \]

VT3 is weaker and captures capacity views, while VT4 is stronger and excludes such views. In this chapter, I examine whether VT3 or VT4 are viable alternatives to VT2, or whether they are \textit{ad hoc} in a way similar to that of Adams’s unwarranted extension of the conception of voluntary control. There is at least one reason to think that VT3 and VT4 may be similarly problematic: the
conditions of direct responsibility for actions and the conditions of direct responsibility for omissions may be substantially different. If so, it is not at all clear that agents can be directly responsible for omissions in the same way and for the same reason that they are responsible for intentional actions.

Ultimately, I conclude that both VT3 and VT4 are problematic as volitionist theses—though for different reasons. I begin with the stronger VT4, arguing that the reasons the volitionist has for focusing on intentional action do not obviously or neatly extend to justification for including intentional omissions. Unless volitionists can offer further reasons justifying direct responsibility for intentional actions that also apply to intentional omissions, VT4 appears to be unjustifiably ad hoc. Then I turn to VT3, which is weaker and broader. I argue that VT3 is not robust enough to ground our responsibility ascriptions and practices, as it makes no appeal to the regard or respect the agent expresses toward others in her behavior. Attempts to add conditions to VT3 bring it substantially closer to nonvolitionism.

3.2 Expressing Who We Are As Agents

I will begin by focusing on VT4, and then I will turn my attention to VT3 at the end of the chapter. To be justified in extending VT2 to VT4 by adding intentional omissions, we must examine what justification there is for thinking agents are directly responsible for their intentional actions and see if that justification extends to intentional omissions. If the reasons why an agent is directly responsible for intentional actions apply equally well to intentional omissions, then prior choice theorists and endorsement theorists can straightforwardly endorse VT4 rather than VT2. Of course, these reasons must apply only to intentional actions and intentional omissions. If these reasons justify more than intentional actions and intentional omissions, then the volitionist thesis would have to include these other items as well. The task, then, is determining what justificatory reasons the volitionist has for focusing on intentional action.

One candidate is that intentional actions express who one is in an important way. By acting intentionally one allows one’s beliefs and desires to manifest in some action. One

39 For instance, Clarke writes that many philosophers hold that there is an important asymmetry between responsibility for actions and responsibility for omissions because they hold that the ability to do otherwise is not required for direct responsibility for actions, but it is required for responsibility for omissions (2014: 135). Clarke cites Fischer (1985-86), Fischer and Ravizza (1991), and van Inwagen (1978) as all subscribing to this asymmetry thesis at some time (though they do not all still subscribe to this thesis).
expresses one’s values and commitments in the world. This point might be put in different ways. One might claim that our intentional actions express the quality of our wills, our true or inner selves, or our moral characters. My intentional action of saving a drowning child shows that I am a caring, selfless agent who recognizes and responds to moral reasons. My intentional action of choosing to insult someone who cuts in line shows that I am an impatient, rash person who desires to hurt others when I have been wronged in some way. The general lesson is that our intentional actions show the world who we are.

Happily for the volitionist, intentional omissions also seem to express our true selves and show the world our commitments, values, beliefs, and desires. When Val intentionally omits to pick up a friend from the airport despite her promise because she simply doesn’t feel like it, Val reveals something morally flawed about herself. She shows that she is selfish and doesn’t take her commitments seriously. When Walt intentionally refrains from punching an antagonist in the face because his antagonist is a fellow moral agent capable of suffering, Walt reveals something morally praiseworthy about himself. He shows that he is patient and that he understands, values, and is responsive to moral reasons. If both Val and Walt can reveal their true selves by intentionally omitting as well as by intentionally acting, the volitionist has found some justification for VT4.

Of course, the volitionist can endorse VT4 only if this justification does not include direct responsibility for any other items. Unhappily for the volitionist, this does not seem to be the case. The nonvolitionist often appeals to this very reasoning to explain why agents are directly responsible for attitudes and various unintentional omissions. These items, no less than intentional actions (and perhaps even more so) express who one is and one’s deep commitments. In fact, some of these items simply are who one is. These items are the attitudes that constitute one’s quality of will. As such, they are intimately tied up with our agency in a way that volitionists who appeal to this justification simply cannot escape.

For example, Angela Smith, a nonvolitionist, writes that what we notice and neglect, what occurs to us, and our nonvoluntary reactions all “have an enormous amount of expressive significance” (2005: 242). Nevertheless, these are not intentional actions or intentional omissions. Smith describes Captain Wybrow, a character from George Eliot’s *Scenes of a Clerical Life*. Wybrow offers Miss Assher, his fiancé, some gelatin in an attempt to be polite. Miss Assher bristles at the invitation, angry that Wybrow has failed to notice after all this time.
that she does not like gelatin. Smith writes, “[Captain Wybrow’s] failure to show any special awareness or appreciation of her own tastes and preferences, in a way that would distinguish her from any other lady who happened to be sitting at the table next to him, suggests to Miss Assher that she does not yet occupy a distinctive place in his overall emotional and evaluative outlook” (243). Wybrow’s failure to notice Miss Assher’s preferences, while not an intentional omission or intentional action, expresses that he is not an agent who cares sufficiently for his fiancé. Similar things can be said for what occurs to us and for some of our unintentional and nonintentional reactions. What we desire, what spontaneously comes to mind, and how we unintentionally react to various situations can all express who we are. Nevertheless, none of these things are intentional actions or intentional omissions.

If the volitionist appeals to the expressive significance of intentional actions and intentional omissions as grounding one’s direct responsibility for those intentional actions and intentional omissions, then the volitionist must also accept that what we notice and neglect, what occurs to us, and our nonvoluntary reactions also qualify for direct responsibility in virtue of this same expressive significance. Appealing only to the expressive significance of intentional actions may explain responsibility for intentional omissions, but it also gets the volitionist much more than she bargained for.

3.3 Exercising Control

The volitionist likely is not concerned with intentional actions only because they express who we are as agents. While this might be relevant to responsibility, the proponent of VT4 might claim that it is not enough. Instead, an agent must exercise some kind of control over an item in order to be directly responsible for it. While we exercise this kind of control over intentional actions, we do not exercise this same kind of control with respect to the things one notices and neglects or with one’s attitudes and emotions. Importantly, the volitionist must argue that we exercise this same kind of control in intentional omissions, so we have good reason to accept VT4 over VT2.

It is tempting at this point to think that we have simply argued in a circle. I claimed that we should abandon talk of voluntary control and focus on intentional actions instead, and now we have returned to talk of control in understanding why attributions of direct responsibility are restricted to intentional actions. But notice that this account of control need not be voluntary
control. After all, there are many different kinds of control. So I think this charge of circularity can be resisted.

I claimed that the volitionist must argue that we exercise the same kind of control in intentional omissions that we do in intentional actions, so VT2 can be extended to VT4. In light of this, I will examine an account of control developed by John Martin Fischer and Mark Ravizza 1998. Although VT4 does not require this account of control, Fischer and Ravizza argue that the same kind of control needed for moral responsibility for actions is needed for moral responsibility for omissions. If Fischer and Ravizza are correct that the responsibility-level control in actions and omissions is the same, this provides some support for the idea that we could add on intentional omissions as in VT4. I will argue that Fischer and Ravizza are wrong, however, so there is no support for the idea that we can add on omissions in this way. In 3.4 I argue further that there is good reason to think that omissions don’t involve an exercise of control at all, and so this justification for holding agents directly responsible for intentional actions does not similarly apply to holding agents directly responsible for intentional omissions.

According to Fischer and Ravizza, an agent S exercises guidance control in acting if and only if (i) S acts from her own mechanism, and (ii) the mechanism is moderately reasons-responsive. The first condition is an ownership condition, and it is to be understood as an historical condition that involves taking responsibility for actions issuing from the mechanism by which one acts. To take responsibility for actions issuing from some mechanism, S must view herself as an agent and an apt target for the reactive attitudes on the basis of some evidence in the appropriate way (Fischer and Ravizza 1998: ch. 8).

The second condition says that the mechanism must be moderately reasons-responsive (1998: ch. 3). A mechanism is moderately reasons-responsive when it is weakly reactive to reasons and regularly receptive to reasons. Reactivity to reasons involves the capacity to translate

---

40 It is not clear whether Fischer and Ravizza intend to restrict their account of control to intentional actions or to actions more generally. In discussing their account, they write, “[T]he agent’s action in the actual sequence must be intentional, that is, appropriately connected to his reasons” (1998: 81). Earlier they claim, “[I]f an agent is to display guidance control, the actual sequence must also exhibit the appropriate sort of connection between reasons and action. So, when an agent has guidance control, we assume that he performs the relevant action intentionally (i.e., for a reason)” (64). Fischer and Ravizza also discuss omissions, but it is similarly not clear that they restrict their account of control to only intentional omissions. While the agents omitting in their cases choose to omit or “deliberately refrain” from doing something, it is not clear that this intentionality is required for their account of responsibility-level control over omissions (126). Whether Fischer and Ravizza intend to restrict their account in this way should not affect my argument. If a different sort of control is required for actions than for omissions, then a different sort of control will be required for intentional actions than is required for intentional omissions, since intentional actions and intentional omissions are a subset of actions and omissions.
reasons into choices and subsequent behavior, while receptivity to reasons involves the capacity to recognize the reasons that exist. For an agent’s mechanism to be weakly reactive to reasons is for the mechanism, when held fixed, to react to some reason to do otherwise in some relevant possible world. According to Fischer and Ravizza, receptivity to reasons must be stronger than this, however. Again holding the mechanism generating one’s action fixed across relevant possible worlds, that mechanism must recognize an understandable pattern of reasons, including how reasons fit together, why one reason is stronger than another, etc. Additionally, at least some of these reasons must be moral reasons. So for Fischer and Ravizza, an agent exercises guidance control in acting only if, when holding the mechanism generating her action fixed, that mechanism is receptive to an understandable pattern of reasons for the action, some of which are moral, and that mechanism generates at least some different action in some relevant possible world on the basis of at least some of those reasons.

To better understand guidance control, consider an action like raising one’s arm. Suppose Gerry is at the doctor’s office and the doctor asks Gerry to raise his arm, and Gerry does so. Assume that the mechanism by which Gerry raises his arm is his own, and so satisfies the first condition. Assume also that this mechanism issues in Gerry’s arm raising because his doctor asks Gerry to raise his arm, and it recognizes this reason as sufficient for issuing in Gerry’s arm raising. The mechanism is receptive to reasons in a patterned way. The mechanism recognizes that there are types of reasons not to issue in Gerry’s raising his arm, and it would fail to generate Gerry’s arm raising on the basis of these reasons if they are strong enough to outweigh the reasons to generate Gerry’s arm raising. For instance, if Gerry would instantly drop dead as a result of raising his arm, the mechanism would recognize this as a reason not to generate Gerry’s arm raising, and Gerry would not raise his arm. If raising his arm would cause an angry dog to attack a child outside the doctor’s office, the mechanism would recognize such a reason and would not generate Gerry’s arm raising. Because Gerry’s mechanism is moderately reasons-responsive, Gerry exercises guidance control in raising his arm at the doctor’s request. Insofar as guidance control is required for responsibility, Gerry is responsible for raising his arm.

Imagine a similar, but importantly different case, in which an agent does not exercise guidance control. Suppose that Gary suffers from spontaneous arm spasms that send his arm straight into the air. When the doctor asks Gary to raise his arm and Gary’s arm shoots into the air, Gary does not exercise guidance control. The mechanism that produces the spasm does not
seem to be receptive to reasons, let alone the same reasons to which Gerry’s mechanism is receptive. Nor is Gary’s mechanism reactive to reasons.\textsuperscript{41} This is a good result: Gary is not responsible for raising his arm when the doctor asks him to, because Gary does not exercise guidance control in raising his arm.

A mechanism may also fail to be receptive to reasons, or reasons of the right sort. In a psychopath, for instance, the mechanism that operates in generating some behavior may fail to be receptive to moral reasons. As a result, the psychopath fails to exercise guidance control with respect to that behavior (84-85). Or one’s operating mechanism may fail to be receptive to an understandable pattern of reasons. If an agent acts from a mechanism that would accept the fact that someone on the ferry is smoking a Gambier pipe as the only reason not to kill all the passengers of the ferry with a saber, that mechanism is not regularly receptive to reasons (84). Consequently, the agent fails to exercise guidance control in killing the passengers.

Given this account of guidance control, it is clear that guidance control is not coextensive with intentional action. Instead, guidance control is more stringent than what is required for intentional action. One might act intentionally even though the action-generating mechanism is not moderately reasons-responsive or even though the agent has not taken ownership of the mechanism. Nevertheless, there are important lessons to be learned from Fischer and Ravizza’s attempt to show that the same sort of control exercised in actions is exercised in omissions. Fischer and Ravizza propose an account of direct responsibility for omissions that parallels their account of responsibility for actions. They claim that guidance control is the same sort of control at issue in omissions. If this is right, this lends some support to the idea that volitionists might be justified in including omissions in their thesis as in VT4.

To understand how Fischer and Ravizza view guidance control as it relates to omissions, we must first distinguish between what they call \textit{simple omissions} and \textit{complex omissions}. According to Fischer and Ravizza, simple omissions are failures to move one’s body in a certain way or hold one’s body still (132). Because simple omissions involve only a lack of bodily movement, Fischer and Ravizza also call simple omissions “bodily omissions.” Some examples of simple omissions are a failure to raise one’s arm or a failure to keep one’s eyes forward.

\textsuperscript{41} There are worries here about how a mechanism is individuated. Is the mechanism that results in Gary’s spasms the same mechanism that would raise Gary’s arm for reasons? This is an important question, but Fischer and Ravizza do not provide an account of mechanism individuation. I simply raise the issue here to set it aside.
According to Fischer and Ravizza, “one has guidance control of one’s failure to do \( A \) (in a case of a bodily omission) just in case one’s actual bodily movement \( B \) (which fully constitutes the omission) issues from one’s own, moderately reasons-responsive mechanism” (133). Just as guidance control is the kind of control required for responsibility for actions, guidance control is the kind of control required for responsibility for omissions. So imagine a third doctor case in which Harry is asked to raise his arm. The mechanism that generates Harry’s arm staying still at his side is moderately reasons-responsive; it is receptive to an understandable pattern of reasons for raising his arm, some of which are moral, and it would raise his arm in response to and because of at least some of these reasons. Yet as it happens, Harry doesn’t raise his arm when the doctor asks him to. Harry is responsible for not raising his arm, according to Fischer and Ravizza, because he has guidance control with respect to this simple omission. The account neatly parallels the account of responsibility for actions.

Fischer and Ravizza recognize that many omissions are not simple omissions, however, but complex omissions. Complex omissions are not fully constituted by bodily movements (133). Instead, they involve a bodily movement and an individual’s being related to some sort of consequence. In this way, complex omissions might be compared to an action plus its consequences. For instance, John’s standing on the shore, not saving a child from drowning is a complex omission. It involves John’s bodily movement (or rather, the way John moves his body—in this case, by holding his body still) and his relation to the consequence of the child drowning. Fischer and Ravizza say that it is useful to think of complex omissions “as the agent’s bringing about relatively finely specified negative consequence-universals” (134). So, for example, when John fails to save a child from drowning, John brings about the negative consequence-universal “that the child is not saved by him” (134).

Responsibility for complex omissions is more complicated than that for simple omissions, because one must be related to the consequence in the appropriate way. Not only must the agent exercise guidance control in failing to move her body in some way, but the process that leads to the consequence-universal must be responsive to this movement. So, for example, suppose that John exercises guidance control as he fails to move his body while standing on the shore. Suppose also that a nearby child drowns. In order for John to be directly responsible for the omission of not saving the drowning child, the fact that the child is not saved by him must be sensitive to the way in which John fails to move his body. To determine whether
John is responsible for this consequence, we examine a range of possible scenarios where John moves his body in some different way. To be responsible, the actual mechanism producing the behavior in this range of scenarios must be reasons-responsive, and there must be some possible scenario in this range where John has reason to move his body in a different way and does so for that reason. But additionally, it must be the case that if John were to move his body in that way, and we hold fixed the non-occurrence of other triggering events that did not actually occur, then the consequence-universal that the child is not saved by him would not obtain (see 135).

Despite the neatness of Fischer and Ravizza’s account, it is far from clear that guidance control is the sort of control exercised in direct responsibility for omissions. Clarke argues that guidance control is neither necessary nor sufficient for responsibility for simple omissions. He also suggests that the account of the sort of control at issue in complex omissions is unclear or incomplete at best, and false at worst. I’ll begin with simple omissions before turning briefly to complex omissions.

To be responsible for a simple omission, according to Fischer and Ravizza, one’s actual movement (or stillness) must issue from a moderately reasons-responsive mechanism. Supposing that the failure to perform some action, A, is the actual movement of one’s body in a certain way B that issues from a mechanism of kind M, what it means for one’s failure to do A to issue from a moderately reasons-responsive mechanism is the following:

It must be the case that, holding the operation of M and the natural laws fixed, there is a suitable range of scenarios in which the agent recognizes sufficient reasons (some of which are moral reasons) to move in some alternative way B*, and in at least one scenario in which the agent has a sufficient reason to move in a way B*, he does so (for that reason), and his doing so would count as his doing A. (Fischer and Ravizza 1998: 133)

There seem to be cases in which one’s failure to do A issues from a moderately reasons-responsive mechanism in this way, and yet this guidance control is not sufficient for an agent’s responsibility for this failure to do A.

---

42 Of course, it is questionable whether guidance control is the right sort of control for responsibility for actions as well. I return to this point below. For important criticisms of Fischer and Ravizza’s account of guidance control with respect to actions, see Mele 2000 and 2006a.
Consider the following case, offered by Randolph Clarke (2014: 125):

*Steel Bars:* Fran is sitting, eyes closed, hands resting in her lap. She had promised to raise her right arm now, but she decided not to bother. Unbeknownst to her (and through no fault of hers), a steel cage has been placed around her upper body. No matter how hard she might try to raise her arm, she wouldn’t be able to do so.⁴³

Clarke claims that Fran is clearly not responsible for not raising her arm: “Fran might be responsible for not trying to raise her arm. But plainly she isn’t responsible for not raising her arm. No matter how hard she might try to do so, she wouldn’t be able to do so, and she’s not responsible for that being so” (125). Yet as *Steel Bars* is presented, she seems to have guidance control with respect to the mechanism that issues in her failure to raise her arm. After all, as Clarke points out, the cage is not part of the mechanism that issues in the positioning of Fran’s arm. Fischer and Ravizza hold that we should focus on the actual sequence of Fran’s behavior, but because the cage is not part of the mechanism, it is irrelevant to the actual sequence of her behavior. If we excise the steel cage from *Steel Bars*, we can clearly see that the mechanism issuing in the stillness of Fran’s arm is her own moderately reasons-responsive mechanism. Fran’s mechanism might recognize sufficient reasons, some of which are moral, to issue in her raising her arm in a suitable range of scenarios. And the mechanism might generate her moving her arm for one of these reasons in at least one of these scenarios, because these scenarios include many in which the cage is absent. But even with the cage in place, the mechanism still counts as reasons-responsive, because the cage is not part of the mechanism.

---

⁴³ Clarke notes that Fran’s failure to raise her arm in *Steel Bars* may not count as an omission—particularly if omitting to *A* requires an ability to *A* (2014: 125). Nevertheless, we can ask whether Fran is responsible for not raising her arm. Presumably, Fischer and Ravizza are interested in answering this question with their account. In fact, they might have no problem referring to Fran’s failure as an omission: “[A]gents seem to omit to do certain things they in fact cannot do. If this is correct, then it is not appropriate to claim that it is a conceptual requirement on omitting to do *A* that one have the power to do *A*” (Fischer and Ravizza 1998: 125, n. 3). I don’t aim to resolve this issue here, but merely to raise it as a potential concern.

Still, one might protest that if Fran’s failure to raise her arm is not an omission, *Steel Bars* changes the topic. The question at hand was whether the same sort of control that grounds responsibility for actions can explain responsibility for omissions. Accordingly, *Steel Bars* can show us nothing about the sort of control that explains responsibility for omissions.

In response, it should be noted that whether Fran’s failure to raise her arm should be understood as an omission or not, this failure is something for which an agent can be responsible or not responsible, and a comprehensive theory of responsibility must account for this. Fischer and Ravizza would very likely agree, and it is for this reason that they do not wish to narrow the conception of omission in the above way.
Of course, one might question Clarke’s judgment that Fran is not responsible for not raising her arm. After all, she decided not to bother to raise her arm on her own, and the cage had nothing to do with her decision. Clarke contends that this conclusion would be rather hard to swallow by presenting an augmented case in which Fran must raise her arm to save a child from death. Fran does not raise her arm, and the child dies (2014: 126, n. 18). It seems that Fran is not responsible for not saving the child, since no matter how hard she might have tried to save the child, she simply could not have saved the child due to the cage surrounding her. If one claims that Fran is responsible for not raising her arm, then one must also claim that Fran is responsible for failing to save the child. It seems fairly clear, however, that Fran simply is not responsible for failing to save the child in virtue of her circumstances, and these same circumstances preclude her being responsible for not raising her arm. Consequently, guidance control is insufficient when it comes to responsibility for such absences.

Clarke offers another counterexample to suggest that guidance control isn’t necessary for responsibility for simple omissions either:

*Spasm*: A doctor has implanted a microchip in my motor cortex, which now activates spontaneously, causing my arm to flail about in a spasm. The doctor is monitoring my thoughts, and if I decide to scratch my head, she’ll see that I’ve so decided, she’ll deactivate the chip, and I’ll carry out my decision. I’m aware that I’d be able to carry out such a decision. I don’t so decide, and the doctor has nothing to do with my not so deciding. (2014: 126-127)

In *Spasm*, Clarke does not decide to scratch his head, and so he does not raise his arm. But Clarke judges himself responsible for not raising his arm to scratch his head in this case (127). After all, he is aware that if he decides to raise his arm, he will be able to raise it, and he actually is able to raise his arm. If he decided to scratch his head, nothing would have prevented him from raising his arm and carrying out that decision. The doctor would have deactivated the chip, and a different mechanism that is moderately reasons-responsive would have taken over. As it happens, however, the mechanism that moves Clarke’s arm spastically is not moderately reasons-responsive; he exercises no control over this mechanism. Given that Fischer and Ravizza are focused on the actual sequence and the actual mechanism, however, the mechanism that
would kick in were Clarke to decide to scratch his head is irrelevant. What matters is the mechanism that is actually causing his arm to move, and this mechanism is certainly not reason-responsive. So *Spasm* shows that guidance control is not necessary for responsibility for simple omissions.

If guidance control is not the sort of control at issue in simple omissions, then it will not be an accurate guide to responsibility in complex omissions. According to Fischer and Ravizza, in order for an agent to be responsible for a complex omission, the movement of her body must be an exercise of guidance control. Additionally, if the agent were to move her body differently, this would be a triggering event that initiated a process such that the consequence would not occur (holding fixed all other triggering events that do not occur such that they do not occur) (135). Unfortunately, Fischer and Ravizza are quite unclear regarding the notion of a triggering event. But the details of their account of responsibility for complex omissions are not as important for our purposes here. If guidance control is not necessary or sufficient for responsibility for simple omissions, then its connection to responsibility for some consequence is irrelevant. An agent must be responsible for some simple omission in order to be responsible for the complex omission. Given that Fischer and Ravizza rely on the problematic notion of guidance control to explain responsibility for simple omissions, and given that guidance control is neither necessary nor sufficient to explain this responsibility, it is of no interest to us here how such simple omissions are connected to consequences.

Fischer and Ravizza intend that guidance control explain why and how agents are responsible for actions, but guidance control cannot explain why and how agents are responsible for simple omissions. In light of the arguments above, it seems that if guidance control is required for responsibility for acting, then a different sort of control is required for responsibility for omitting. While the sort of control required for intentional actions might be different from the sort of control required for responsibility for actions, the fact that responsibility-level control

---

44 Clarke offers several illustrations of the way in which this notion is problematic (2014: 127-132).

45 There are also problems in understanding omissions as simple or complex, as Fischer and Ravizza do. As Clarke points out, omissions of mental actions are neither bodily omissions nor complex omissions (2014: 21). Mele makes a similar point (2000: 451-452). In fact, Fischer and Ravizza’s view on the nature of omissions seems deeply problematic. According to their view of simple omissions, “the way one actually moves one’s body (where this can include simply keeping the body still) ‘fully constitutes’ the omission” (1998: 132). Clarke refers to this view of omissions as the constitution view and argues convincingly that it is false. Most simple omissions are absences and are not constituted by bodily movements (2012, 2014: ch. 1). I say more about the nature of omissions and how this might be relevant to the justification of VT4 in section 3.4 below.
involved in actions and omissions is not the same undercuts support for the idea that actions and omissions involve the same kind of control. There is, at least, no positive case yet to be made that the same sort of control that is required for intentional actions is required for intentional omissions, and consequently there is no positive reason to accept VT4 over VT2.

One might think that a positive case can be made for VT4 by turning to a different account of control. Fischer and Ravizza hold that when an agent acts intentionally and is responsible for her action, the reason she is responsible for so acting is because she exercises guidance control. Yet one might reject this supposition, arguing that guidance control is not the right sort of control required for responsibility. Perhaps a different account of responsibility-level control could give the volitionist the right result such that when one is responsible for one’s conduct, this is in virtue of exercising the same control in both acting and omitting.

There is good reason to think that when one omits (where this omission is an absence), one does not exercise the same type of control that one exercises when one acts intentionally regardless of the account of control one favors. It doesn’t matter what sort of control one claims is exercised in intentional actions; no such control will be exercised in intentional omissions. Instead, an omission involves a capacity to exercise some type of control. To develop this point, it will be beneficial to explore a causal theory of action. If intentional actions and intentional omissions both involve an exercise of the same type of control, the way in which an intention brings about an action or an omission ought to look quite similar. As I will argue in 3.4, this does not seem to be the case.

3.4 The Causal Theory of Action

On the surface, intentional actions and intentional omissions seem to enjoy many similarities. When an agent intentionally acts or omits, she acts or omits in that way in virtue of some intention. Both seem to be caused by intentions. Both can result from deliberation and choice. Both can be done for reasons. Both can require effort or skill.

Let’s first focus on the way intentions bring about some action. A causal theory of intentional action enjoys wide support. According to causal theories of intentional action,
Typically, this mental state will be an intention, but an event, such as the acquisition of an intention, might also suffice to cause the agent’s behavior (Mele and Moser 1994).

Because actions might be deviantly caused by a mental state or event, causalists about action include a non-deviance clause. For an action to be caused “in the right way” is for it to be non-deviantly caused by the mental state. To take a familiar example, wanting to make his climb easier, a climber might intend to let go of the rope supporting his companion. The climber might be so shocked that he has this bad intention that he loses his grip on the rope. Although the climber’s intention to let go of the rope caused him to release his grip, the intention didn’t cause the release in the right way. That is, the climber did not intentionally let go of the rope, even though his intention caused him to let go of the rope. While causal deviance is a challenge for the causalist, some have proposed solutions (e.g., Mele 2003: 51-63). Since we need not be concerned with the success of CA, I do not evaluate these solutions here.

Given that a volitionist might favor CA, one might wonder what is to stop the volitionist from favoring a causal theory of intentional omissions, such that both are united under the causalist banner. Perhaps, for instance, one could claim:

\[ CO: \text{Necessarily, } S \text{ intentionally omits only if } S^\prime \text{ s omission is caused in the right way by an appropriate sort of mental state or event.} \]

If so, then the volitionist might be able to bolster support for VT4. If the ways actions and omissions are brought about by intentions are similar, this suggests that whatever control is exercised in bringing about the one is also exercised in bringing about the other.

Given the nature of omissions, this move appears rather dubious. Intentions cause actions, and actions in turn have causal power. Recall that many omissions are nothing at all—

---

\[^{46}\text{For causal theories of intentional action, see Mele 1992 and 2003, Mele and Moser 1994, and Clarke 2010b. In light of concerns about mental causation, Mele holds that the neural realizers of the relevant mental states might be sufficient to play the relevant causal role (1992: ch. 2). A more precise version of CA may disjunctively include these neural realizers, but for the sake of simplicity I omit Mele’s more careful formulation here and in what follows.}\]

\[^{47}\text{This example is from Davidson (1980: 79).}\]
they are absences of actions, rather than, say, negative actions or positive actions described negatively (Clarke 2014: ch. 1). If many omissions are absences, to adopt CO would require accepting also that absences can be effects of intentions. Because our actions also have causal power, a strict parallel of omissions would also require accepting that omissions can be causes. A causal theory of omissions seems to require a defense of how absences of action can be causes and effects.

There is much that counts against a successful defense of this claim (Beebee 2004; Clarke 2014: 54-58). One powerful consideration against absence causation, for instance, is that in order to have causal powers, something must have some properties to ground these causal powers. But absences aren’t things, so they certainly aren’t things with properties. Consequently, absences can’t be causes or effects (Clarke 2014: 55). If many omissions are absences, and absences can’t be causes or effects, many omissions aren’t even plausible candidates for a causalist theory like CO.

While CO may not be a viable candidate, something like CO may suffice. For instance, Clarke argues that “absences can result and have results even if no absences are causal effects or causes” (2014: 51). Instead of focusing on the absences, we can focus on the positive events—the actual processes that took place and counterfactual processes that would have taken place if things were different. Suppose Eve intentionally omits to walk her friend’s dog as promised because she would rather watch Netflix. It is true, we can suppose, that if Eve had intended to walk her friend’s dog, she would have walked the dog. It is also true that as a result of her desire to watch Netflix, Eve omitted to walk her friend’s dog. This may be causation (if absences can be caused), but it need not be. Similar things can be said about results or consequences of omissions.

If omissions can result and have results, the problem of absence causation might be avoided. In the case of intentional omissions, the intention need not cause an absence, but it might merely need to cause some of the agent’s thoughts or conduct such that the omission can be said to result or have results. This seems to be Clarke’s suggestion for the causalist about intentional omissions:

---

48 I assume here that an absence is, in fact, nothing.
A standard causal theory of action holds that, in a case of intentional action, an intention with relevant content must appropriately cause certain agent-involving events. It seems that in a case of intentional omission or refraining, likewise, an intention with relevant content must play an appropriate causal role with respect to subsequent agent-involving events. (2014: 81)

So what counts as relevant content?

According to Mele 1992, the representational content of an intention to act is an action plan. Parallel to this, Clarke claims that the content of an intention not to \( A \) is a plan not-to-\( A \) (2014: 71). To intentionally omit, one might have an intention not to \( A \), and thereby be committed to the plan not-to-\( A \). But one need not have the intention not to \( A \) in order to intentionally omit to \( A \). For instance, a smoker who knows that he is weak-willed and suspects that he will not be able to quit smoking for a week might cautiously intend only to try not to smoke for a week. If he succeeds in not smoking for a week, this is an intentional omission (66). So one’s omission might count as intentional not only if one has an intention with the plan not-to-\( A \), but also in cases where one has the intention to try not-to-\( A \). Perhaps other content of an intention might count as relevant, but with Clarke, I assume this is a vague matter (71).

Provided that the intention not to \( A \) (or to try not to \( A \)) plays the appropriate causal role with respect to the relevant agent-involving events, an omission will count as intentional. Just as CA rules out causal deviance from intentional action, Clarke wants an intention’s playing the appropriate causal role to rule out deviant causation with respect to omissions as well. When Ann forms an intention not to pick up Bob at the airport, her omission does not count as intentional if this intention immediately triggers an episode of narcolepsy that results in her failing to pick up Bob (80). While Clarke does not give us many details about how to rule out causal deviance, he does claim that one’s relevant intention must cause what it does “[s]o that one’s control over whether one does the thing in question isn’t lost” (81). Notice that this is control over something, which suggests an ability to control rather than an exercise of control.

---

50 See Mele (1992: chs. 8 and 9).
51 Joshua Shepherd 2014 argues that Clarke is not entitled to this non-deviance claim given the work that his view currently does: “[i]t is difficult to believe that any of an agent’s relevant mental states contribute essentially to her control over subsequent thought or behavior. An agent’s control plausibly stems from much more than any one intention of the agent” (12). If the mere causal relation of the intention with some behavior is not sufficient to secure the relevant notion of control, Shepherd claims, then Clarke’s view is at best incomplete and at worst false.
Clarke’s causalist account of intentional omissions does not require absence causation. One’s not doing \textit{A} might result from an intention not to \textit{A}. One’s intention not to \textit{A} might cause some behavior that is incompatible with \textit{A}-ing. And perhaps one would have \textit{A}-ed had one not formed the intention not to \textit{A}. We need not make any contentious claims that an intention has caused an absence on this picture. Yet we can still understand intentions as serving an important causal role with respect to intentional omissions.

Even if Clarke’s causalist account is successful, this does not yet mean that a prior choice theorist or endorsement theorist can embrace VT4 over VT2. What this exploration has revealed is that the way intentions are causally related to actions seems significantly different from the way that intentions are related to omissions. Intentions to \textit{A} often sustain and guide one’s \textit{A}-ing, but intentions not to \textit{A} need not guide or sustain any particular behavior. In order for the intention in an intentional action to be causally related to the action, the intention has to cause the particular action. But in order for the intention in an intentional omission to be related to the omission in the right way, the intention can cause any sort of behavior or be causally related to any sort of behavior other than the behavior in question. Clarke puts the point this way:

Generally, there’s much less that an intention not to act must do in order to secure its satisfaction than is required of an intention to act. Where \textit{A}-ing is performing some ‘positive action,’ \textit{A}-ing generally requires that one’s behavior fall within some fairly narrow range of possible behaviors, while not \textit{A}-ing generally requires only that one’s behavior remain outside—anywhere outside—some quite narrow range. Hence, the work demands on intentions not to act are usually comparatively light. They can generally secure their satisfaction by doing relatively little. (2014: 79)

While intentions must do work in both intentional actions and intentional omissions, the amount of work they do in each is quite different, and the way in which the intentions do their causal work is substantially different.

This leads Shepherd to propose the following causalist account of intentional omissions: “Necessarily, an agent \textit{J} intentionally omits to \textit{A} only if an intention of \textit{J}’s with relevant content (or the intention’s acquisition) non-deviantly causes in \textit{J} a disposition not to \textit{A} (where non-deviant causation here involves the intention’s making changes to \textit{J}’s cognitive and motivational systems that are coherently related to the intention)” (18). I am not interested in comparing Shepherd’s causalist account of intentional omissions to that of Clarke’s or evaluating which is better. If Clarke’s view fails, perhaps Shepherd’s dispositional account can do the relevant work. In footnote 52 I consider how my comments below might apply to Shepherd’s view.
Consider an example. In order to intentionally play Rachmaninoff’s “Prelude in C# Minor,” I must form the intention to sit down at the piano and press a specific series of keys in a specific sort of way. This action requires a great deal of skill and effort, and my intention to play the piece causes and sustains my playing. In order to intentionally play the piece, I must perform a very specific series of actions. In order to intentionally omit or refrain from playing “Prelude in C# Minor,” I could do any number of things. I may sit and stare straight ahead, I may continue typing at the computer, or I may skillfully play a different piece on the piano. I may intentionally omit to play the piece without intending that any of these specific alternatives not be playing the piece. Although intentions are causally related to both intentional actions and intentional omissions, it is not at all clear that they are related in the same way or even in a way similar enough to draw the parallels that an advocate of VT4 wants to draw.  

In fact, the sort of control that is at issue in intentional omissions seems to be a different sort from that at issue in intentional actions. As Clarke says, “Unless an intentional omission or refraining is itself an action, it isn’t an exercise of the kind of control that we exercise in intentional action” (2014: 86). One does not perform some action, employ certain skills, or otherwise exercise some ability in many of these omissions that are simply absences of actions. Rather, one has an ability to perform some action or employ some skills and simply refrains from exercising that ability. While intentional action involves an exercise of control, these omissions involve no such exercise of control. The result is that the motivation for endorsing VT4 over VT2 is severely undercut.

My conclusion here is relatively modest and tentative. There may be other justifications for VT4 open to the volitionist, but the volitionist cannot simply take this for granted. She must do substantial and difficult work to justify VT4 over VT2. So I tentatively conclude that the volitionist cannot endorse VT4 without appealing to other sorts of justifications that either let in too much or that are ad hoc.

At this point, the prior choice or endorsement theorist might protest that I have set her up to fail. I reinterpreted the volitionist thesis of VT1 as VT2, which explicitly rules out omissions.  

---

52 Even on Shepherd’s account involving dispositions, a similar problem arises. One need not be disposed to do anything in particular when one is disposed not to A, which is quite different from the case of intentional action. Perhaps, then, this difference points to a deeper gap between intentional actions and intentional omissions. Even if intentions are causally important for both, if the way in which they are causally important is different, this difference may be relevant to responsibility, and may have a bearing on whether the volitionist is justified in endorsing VT4 over VT2.
Then, I showed that omissions cannot be added to VT2 as in VT4. If omissions are excluded from this volitionist thesis, one might think, the blame lies with me rather than the prior choice theorist or endorsement theorist. But this is too hasty. If this point about the exercise of control is correct, VT1a also excludes these intentional omissions, so VT2 is no worse than VT1a. Although VT2 does not include omissions that are absences of action, neither can VT1a. Yet VT2 is, I contend, clearer than VT1a, and therefore an improvement.

3.5 Capacity Views

While VT4 fails, the broader and weaker VT3 is still a live option. The problem with endorsing VT4 is that it does not seem that agents exercise the same type of control in acting intentionally as they do in intentionally omitting. Capacity views remain untouched by this concern, however, as capacity views do not require the exercise of control. Instead, they require the capacity to exercise control with respect to some item. If omissions don’t involve an exercise of the same type of control as intentional actions, it is still open that they involve the capacity to exercise the same type of control as intentional actions. This provides some motivation for capacity views, which are poised to capture many more items under the scope of direct responsibility than prior choice or endorsement views.

While capacity views have earned less attention in the literature on volitionism and nonvolitionism, some writers have suggested capacity views in an effort to account for responsibility for certain omissions, including negligence:

[T]here is a culpability of unexercised capacity in addition to there being a culpability of choice; i.e., one can justly be blamed for wrongful failure to exercise a capacity not to have done a wrongful act, just as one can justly be blamed for having chosen to do such an act. Indeed, as Hart and others have urged, one can view the culpability of choice as being but a special case of the culpability of unexercised capacity: if one adverted to the likely consequences of one’s action, and if one nonetheless chose to go forward, one must have had the capacity not to have so acted (because one had a capacity not to have so chosen). (Moore and Hurd 2011: 151)

---

53 Moore and Hurd are referencing Hart 1968 here. Importantly, they are not endorsing Hart’s view, but merely explaining it.
This suggestion focuses on choice, which as we have seen, is a typical rough gloss on voluntary control. The idea is that, while choosing is morally relevant, it is relevant because it is the exercise of a broader morally relevant capacity: the *capacity* to choose. One need not exercise that capacity in order to be directly responsible. So the focus should be on the capacity rather than the exercise of the capacity.

Consider an example from Randolph Clarke (2014: 14). Seeing a sign that reads “Wet paint: Do not touch,” Clarke does not touch the object. An advocate of VT2 might have a difficult time explaining why Clarke is responsible for not touching the freshly painted object. His refraining was not an intentional action, because it wasn’t an action at all. But, we can suppose, he doesn’t choose not to touch the object either, as he never deliberates about what to do. He simply refrains from touching it. In an intentional action, Clarke exercises some kind of control. In omitting from intentionally acting, as he does here, Clarke does not exercise control, but simply has the capacity to exercise that control. This broad capacity is what VT3 relies upon for ascriptions of direct responsibility.

I have already argued in chapter 2 that VT3 is broader and can capture direct responsibility for far more items than both VT2 and VT4. It even looks like VT3 is prepared to explain how Angela is directly responsible for not keeping her promise in *Birthday*. Angela had the capacity to keep her promise. Not calling her friend Dave is an omission of an intentional action and a violation of some norm or obligations she incurred by promising. Consequently, not calling Dave is something for which she is directly responsible on VT3.

Given the power of VT3 and the apparent difficulties with VT4, why not simply accept VT3? First, VT3 seems unable to satisfactorily explain *why* an agent is responsible for omissions. Consider our reasons for blaming an agent for an omission of an intentional action. What we cite is not simply the capacity for acting, the violation of some norm, or even the combination of these. We recognize that agents might violate some norm in failing to exercise their capacity in some way or other and yet still think that such agents aren’t responsible for their omissions. The reason, I suggest, is that we care about some further features of the agent. We care about what her actions and omissions signify about her. Does the agent express some disrespect or reflect some lack of respect in her action or omission?54 If she doesn’t, then it

---

54 I use “express” to signify an attitude being positively displayed in some agent’s behavior, and “reflect” to signify an attitude or lack of attitude that one might infer from some agent’s omission. Although I focus on blame here and
seems inappropriate to hold her responsible. Prior choice theorists attempt to explain this expression via some exercise of control. But capacity theorists might seem content to allow that an agent can deserve blame, for example, if she violates some norm and had the capacity not to do so. As far as I can see, this is simply too weak to justify attributions of responsibility.\textsuperscript{55}

Of course, VT3 is simply a necessary condition, and this objection treats the capacity theorist as if she is offering a sufficient condition on responsibility. The capacity theorist is free to add a further necessary condition about the agent’s attitudes, regard, or quality of will. But perhaps she might not think she needs to. Perhaps, she might argue, this expression or reflection of an attitude comes for free. Whenever an agent behaves in some way, she expresses some attitude, and whenever she fails to exercise some capacity, she reflects some other attitude or some lack of attitude. Neil Levy offers a compelling argument that this not the case, and I examine this argument in 7.3.2. Sometimes an agent simply fails to act and this failure does not reflect any attitude or lack of attitude. The condition does not, it seems, come for free.

The capacity theorist can simply add a further necessary condition in conjunction with VT3, however, such as:

\textit{AER}: An agent is directly responsible for some intentional action or omission of an intentional action only if that intentional action or omission of an intentional action expresses or reflects the agent’s attitudes or values.

A capacity view that also accepts AER (or something like it) is, I think, a formidable volitionist view. But AER requires spelling out what this expression or reflection comes to and why it is significant for responsibility. It seems to me that the best story, as I argue in chapter 5, is a nonvolitionist one. That story will require an appeal to some sort of normative rational connection between an agent’s attitudes and values. As we will see in chapter 5, AER brings the capacity theorist quite close to nonvolitionism. The capacity theorist who accepts this line would take one of the most important and fundamental conditions the nonvolitionist defends and adopt

\textsuperscript{55} I return to this point in 5.4 when discussing nonvolitionist views. Certain nonvolitionist views focus heavily on what the agent’s actions or omissions express or reflect about her, and this focus is the important piece that capacity views lack.
it to explain responsibility. In fact, it seems this condition is needed to explain why agents are responsible in the first place, and so this condition, rather than the capacity condition in VT3, does the heavy lifting in explaining why agents are responsible for omissions.

The capacity theorist might accept that their view comes close to nonvolitionism—far closer than a prior choice view does. Nevertheless, the capacity theorist appeals to intentional action in VT3, which sets capacity views apart from nonvolitionist views. A condition like VT3 might be in place for the purpose of fairness. It isn’t fair, the capacity theorist might argue, to hold agents responsible for something if the agent lacks the capacity to exercise control over that item. I address this sort of concern in 5.3.2, ultimately arguing that it is not unfair to hold agents responsible in this way. If this argument works, then VT3 is not a necessary condition on responsibility. If AER is the condition doing the work in explaining why agents are responsible, and if VT3 is not required after all, then we should favor a nonvolitionist view.

If that argument fails, however, I think that capacity views are the best alternative to nonvolitionist views. Capacity views that adopt a condition like AER can account for many of our ordinary responsibility ascriptions and practices, as can nonvolitionist views. Ultimately, capacity theorists who favor AER and nonvolitionists might do better to focus on the ways in which they converge and unite against more demanding volitionists, like prior choice theorists.56

VT3 might be an improvement over VT2 in that VT3 includes more items for which agents are responsible. But in order to explain why agents are responsible for these further items, capacity theorists are faced with a choice. If they hold that VT3 is sufficient for responsibility, they lack any grounding that can adequately explain an agent’s responsibility for these omissions. If they add a further condition, this condition will likely have to appeal to an agent’s regard (or lack thereof) for other agents. In doing so, capacity theorists concede much to nonvolitionists and significantly distance themselves from many traditional volitionists, as this further condition will bear the burden of explaining an agent’s responsibility.

56 This idea is similar in spirit to Alfred Mele’s agnosticism regarding whether free action and determinism are compatible. Mele leaves open whether determinism and free action are compatible, but argues against those who attempt to show that agents are not free at all (1995, 2006b). Similarly, capacity theorists and nonvolitionists might do well to unite against prior choice theorists, who seem forced to hold that agents cannot be directly responsible for various omissions of intentional actions.
CHAPTER 4

VOLITIONISM AND DERIVATIVE RESPONSIBILITY

4.1 Introduction

VT3 alone offers little to ground responsibility and VT4 seems unjustified. Prior choice theorists will likely reject VT3, but if VT4 is unjustified, then VT2 seems to be the most defensible thesis for such volitionists. VT2 addresses only intentional actions, however, and we hold agents responsible for a great many items that are not intentional actions, including omissions, act-consequences, character traits, and even some unintentional or nonintentional actions. Consider the following case:

*DUI*: Brad is at a party after a harrowing week. He knows that he must drive home after the party, and he knows that after three beers he will be too buzzed to safely drive home. Nevertheless, Brad decides to indulge and has six beers. On the drive home, as a result of his diminished capacity to operate his car after six beers, Brad swerves his car and crashes into a family sedan, seriously injuring three passengers.

Brad does not intentionally crash into the sedan, so his crashing is not an intentional action. Likewise, the consequence of the crash—namely, that three people were seriously injured—is not an intentional action. According to VT2, Brad is not directly responsible either for crashing into the sedan or for the fact that three people are severely injured. Nevertheless, it is plausible that Brad is responsible for both this unintentional action and this consequence. Cases like *DUI* reflect the need for volitionists who endorse VT2 to explain how agents can be derivatively responsible for items that are not intentional actions.

In this chapter, I explore several conditions on derivative responsibility that proponents of VT2 might endorse. If proponents of VT2 can offer a way to derivatively capture many items for which we hold agents responsible, then perhaps this type of volitionist can offer a theory of moral responsibility that matches our ordinary responsibility ascriptions and practices, and the
worry of chapters 2 and 3 that VT2 fails to capture our responsibility for many items that it should will be unwarranted. In section 4.2, I explore a common technique, called *tracing*, which has been used to explain how agents like Brad in *DUI* can be responsible for crashing into the sedan and for seriously injuring three people. I then propose three possible sets of conditions that a proponent of VT2 might offer to explain derivative responsibility. In 4.3-4.5, I explore these conditions and their drawbacks. Finally, in 4.6, I argue that the shortcomings of these conditions recommend that we take seriously nonvolitionist views to see whether they can better explain our responsibility ascriptions and practices.

### 4.2 Derivative Responsibility and Tracing

The notion of intentional action has built into it some degree of awareness. As I explained in chapter 2, an agent performs a certain type of action $A$ intentionally only if she has some suitable representation to guide the $A$-ing (Mele and Moser 1994: 43). This representation is an *action plan* (42). One need not be constantly aware of this action plan in order to act intentionally, but there must be some sort of representation that guides the action, and the action plan fills this role.

When an agent performs an action intentionally, a volitionist might argue that the agent’s responsibility for the action is explained in part by its representation in the action plan. The action plan ties the action to the agent. Yet in *DUI*, we can suppose that it was no part of Brad’s action plan that he crash into a sedan. Brad’s crashing into the sedan was unintentional for this reason. Still it seems that Brad *is* responsible for crashing into the sedan. This suggests that one’s responsibility for some outcome (where that outcome may be an action, omission, consequence, character trait, or attitude) need not be anchored in the precise moment of acting, or even in the moment right before the outcome occurs. Instead, one’s responsibility can be anchored in “either prior decisions to act [as in a drunk driver case], or in the acquisition of dispositions, habits, or the self” (Vargas 2005: 271). Although Brad’s crashing into the sedan was not part of his action plan, it was something that Brad might have foreseen resulting from his action plan of drinking in excess at the party knowing that he must drive home. We can *trace* Brad’s responsibility for crashing into the sedan to some previous time at which he (i) performed an intentional action for which he was responsible and which led to his crashing into the sedan and (ii) foresaw or was aware that his action might bring about such a result when he performed the action. Theorists can
appeal to this tracing technique generally to explain how agents can be responsible for some outcome, even when that outcome was not an intentional action for which the agent is directly responsible.

Proponents of VT2 who favor a tracing strategy will presumably require that the following two necessary conditions on derivative responsibility:

(1) An agent, $S$, is derivatively responsible for some outcome, $O$, only if $S$ is responsible for some previous intentional action, $A$.$^{57}$

(2) $S$ is derivatively responsible for $O$ only if $O$ is a result of $A$.

Presumably, agents must stand in some epistemic relationship with the outcome as well as with at least some salient features of that outcome that might contribute to its moral status. Volitionists have their pick of how this epistemic status should be understood, but the following necessary conditions seem to be the most plausible and salient candidates:

**AF**: $S$ is derivatively responsible only for outcomes of $S$'s intentional actions that $S$ foresaw would or might occur when $S$ intentionally acted.

**CHF**: $S$ is derivatively responsible only for outcomes of $S$’s intentional actions that $S$ foresaw or could have foreseen would or might occur when $S$ intentionally acted.

**SHF**: $S$ is derivatively responsible only for outcomes of $S$’s intentional actions that $S$ foresaw or should have foreseen would or might occur when $S$ intentionally acted.

---

$^{57}$ I leave open here whether $S$ must be directly responsible or derivatively responsible for $A$. Presumably, this responsibility will be direct, as direct responsibility will provide the appropriate grounding for the derivative responsibility. But if there can be infinitely long chains of derivative responsibility, the responsibility in (1) should not be understood as direct responsibility. Gideon Rosen raises the possibility that “responsibility is a vague matter, and that there can be derivative responsibility without original responsibility for the same reason that there can be chickens even if there was no original chicken” (2008: 299). This is an important possibility, especially given the unclear way in which children eventually become responsible adults. As a result, I leave open that this responsibility could be direct or derivative in (1), but will often assume that responsibility is direct.
As I understand the tracing theorist, one of AF, CHF, or SHF can be added to (1) and (2) to produce jointly sufficient conditions for derivative responsibility. As a result, although AF, CHF, and SHF are necessary conditions, they can be criticized for not being strong enough because no other further conditions are to be added to explain derivative responsibility. If one thinks that SHF is necessary for derivative responsibility, one would not offer the weaker CHF, as SHF arguably entails CHF. Similarly, it would be misleading for one to offer CHF as a necessary condition if one thinks that SHF is required. Consequently, I understand volitionists as offering these necessary conditions with an eye toward sufficiency.

It is important to see that, although S’s intentional action must be prior to some outcome, the action might be fairly far-removed from the outcome. For instance, A may lead to B, which may lead to C, and so on until we reach some outcome. The chain of derivative responsibility may be quite long. Of course, in order for S to be responsible for this outcome, it must be the case that S satisfies AF, CHF, or SHF. S will be derivatively responsible for some outcome at the end of a long chain only if S foresaw, could have foreseen, or should have foreseen that outcome at the time of action (depending on which condition one favors).

Volitionists might disagree about whether AF, CHF, or SHF is the best epistemic condition on derivative responsibility. In the following sections I explore the justifications for each condition as well as their potential difficulties.

4.3 Actual Foresight

AF seems the strongest condition on derivative responsibility, yet it also seems the easiest to defend. If the agent actually foresees some outcome that might result from performing some intentional action, there is a clear connection between the outcome and the agent performing the
action. The agent who performs some intentional action foreseeing that the action will have a particularly negative outcome expresses a particularly objectionable quality of will, we might say, in performing the action. This close connection between the agent and the outcome can explain why the agent is responsible for the outcome, as we simply trace the outcome to some previous point at which the agent performed an intentional action with foresight of the outcome.

What must the agent foresee to satisfy AF? The future is often uncertain. If AF required that agents must foresee with certainty the events that actually take place, responsibility would be rather limited. While agents can sometimes foresee what will actually happen, we do not think that this is required for responsibility. Instead, they must foresee only what would or might happen. And this is precisely what AF requires: $S$ is responsible only for outcomes that $S$ foresaw would or might occur.

But one might reasonably wonder how to understand the “would or might” in AF (as well as in CHF and SHF). This is a difficult question, and there likely is no clear-cut answer. I suspect the volitionist will appeal to the vagueness of our responsibility practices and ascriptions, and AF ought to match that vagueness. Nevertheless, it seems plausible that the bare possibility of some outcome $O$ occurring given some intentional action $A$ is too weak of a requirement. An agent’s intentional action can lead to any number of possible outcomes, many of which are remote. Suppose that whenever I use the microwave, there is the very slight possibility (say, 0.0001%) that it will surge and overload the circuit, cutting power to a variety of devices in the house. My intentional action of using the microwave might cause a surge that causes my friend to lose progress in a video game he is playing. Suppose that I even foresee that this might happen, despite the remote possibility, because I have a vivid imagination. Nevertheless, should I use the microwave and cause a circuit overload, we do not think that I am responsible for the outcome of my friend’s progress being lost. There is simply too much chance involved to take this outcome as something for which I am responsible.

One might, then, think that “would or might” should be understood as “$O$ will probably occur given $A$.” But this requirement is too strong. We may hold an agent responsible for some outcome even if the probability of the outcome occurring given the agent’s prior action was less than 0.5. Instead, I suggest that how likely $O$ must be given $A$ is related to the amount of harm risked in performing $A$. For example, as a prank, Gwen may place a large plastic tarantula on the dashboard of Hank’s car, foreseeing that Hank will or might be startled as he drives. Even if
Hank and Gwen live on a low-traffic road where it is not probable that Hank will hit anyone if he is startled while driving, the amount of harm Gwen risks in placing the fake tarantula on the dashboard is great. Should Hank be startled while he drives so that he swerves into an unlikely pedestrian, for instance, it is no excuse for Gwen to claim that the probability of Hank hitting the pedestrian as a result of her prank was less than 0.5. In this case, the outcome is quite bad, and the reasons for Gwen putting the tarantula on the dashboard are not particularly weighty. Supposing that Gwen was directly responsible for her intentional action of placing the tarantula on Hank’s dashboard, and supposing that she foresaw that placing the fake tarantula on the dashboard might lead to someone being harmed, it seems that Gwen is plausibly derivatively responsible for this outcome.

One might object to AF because actual foresight is too strong of a requirement for derivative responsibility. Consider Manuel Vargas’s case of Jeff the Jerk:

*Jeff the Jerk:* Jeff, a middle-aged manager of a company, must fire some employees as part of a company downsizing in an abysmal economy. Though he does not intentionally choose to fire a group of long-time employees in an insensitive way, Jeff does so because he is an unreflective and habitual jerk. Precisely because he is such a jerk, Jeff is unaware that he is doing anything wrong when he fires the employees so rudely. (adapted from 2005: 271)

Vargas judges that Jeff is clearly responsible for firing his employees in the way that he does, and yet it does not seem that Jeff is directly responsible for the *manner* in which he fires his employees. While the firing of the employees was an intentional action, Jeff was unaware of the way in which he was firing the employees. Vargas’s thought is that for Jeff to be responsible for the way in which he fires the employees, he must be responsible for being a jerk. If Jeff is responsible for being a jerk, then his responsibility for this character trait can carry through to its manifestation in the firing of the employees.

Unfortunately, says Vargas, we cannot trace Jeff’s responsibility back to some suitable prior time. Vargas tells us that when Jeff was 15, he decided to take steps to become more like a jerk in an effort to be successful with the opposite sex. His efforts to become more jerk-like were successful. Of course, “at no point during the process of becoming a jerk, and certainly at no
point before he undertook the program of becoming jerk-like, did Jeff even conceive that his plan for personal improvement would include in its outcomes that he would some day lay off employees in a despicable fashion” (2005: 276). Given Jeff’s age and context, he simply did not foresee this consequence.

If Jeff did not foresee this outcome at his time of action, however, then according to AF, Jeff is not responsible for the outcome. This point generalizes. Consider Brad in DUI. We can assume that Brad did not foresee that he would or might hit some sedan or injure members of some family when he drove home after excessive drinking. Foresight is rarely, if ever, that specific. If this objection is on target, AF would leave few agents derivatively responsible for anything.

A proponent of AF can reasonably resist this line of objection. John Martin Fischer and Neal Tognazzini argue that outcomes can be specified in a range of ways. They offer three ways to specify the outcomes in Jeff’s case (2009: 537):

**Outcome 1:** Jeff fires *those* employees who work for *that* company on *that* precise day in *that* precise manner.

**Outcome 2:** Jeff fires some of his employees at some company or other at some point in the future in a despicable manner as a result of his jerky character.

**Outcome 3:** Jeff treats some people poorly at some point in the future as a result of his jerky character.

While Outcome 1 is likely impossible for 15-year-old Jeff to foresee, Outcome 3 is not. In fact, it is quite plausible that young Jeff actually foresee this outcome. Fischer and Tognazzini suggest that, given this plausibility, we should understand the outcomes in AF broadly, as in Outcome 3. In Brad’s case, he might only need to see that he would or might harm someone in the future as a result of his drunk driving. There is no need to assume that agents must foresee events so specifically, according to Fisher and Tognazzini.

While this strategy of broadly specifying an outcome seems initially plausible, one might sensibly press further. It is not good enough to gesture at some general outcome in understanding
AF. If the outcome is understood too narrowly, agents will be responsible far less often than we ordinarily think they are, as agents plausibly do not foresee many of these outcomes with a high degree of specificity. If the outcome is understood too broadly, however, agents will be responsible too often (supposing that AF, (1), and (2) are jointly sufficient conditions on derivative responsibility). Agents plausibly do foresee that very general outcomes would or might occur as a result of their intentional actions. The proponent of AF needs a way to specify outcomes in AF that matches our responsibility ascriptions, not including too much or too little.

Fischer and Tognazzini address this concern when they advocate specifying outcomes more broadly:

One might be worried that if we are allowed to specify outcomes broadly for tracing purposes, then most people will turn out to be morally responsible for too much. After all, everyone ought to be able to foresee that they might inadvertently offend someone at some point in their lives! While we are inclined to think that our approach does allow for moral responsibility in such contexts, we do not think that this is a problem, since it may well be that you are morally responsible for inadvertently offending someone even though you are not blameworthy for doing so. What sort of epistemic requirement is a condition for blameworthiness is an important and interesting question that we do not take up here. (2009: 554 n. 16)

Fischer and Tognazzini suggest that agents are actually responsible for quite a lot, yet they attempt to make their position more palatable by claiming that there are different epistemic conditions on blameworthiness than there are on responsibility. As I argue in Fritz 2014, separating responsibility for wrongdoing and blameworthiness in this way is untenable. If my arguments there are on target, Fischer and Tognazzini cannot argue that an agent is responsible for some negative item and yet not blameworthy for the item to make their position more palatable. Instead, they must specify an outcome in a more fine-grained way, but not so fine-grained that they fall prey to Vargas’s objection. Admittedly, there may be some way to specify

---

60 If one rejects this supposition, one might propose specifying the outcomes broadly to be cautious, since AF is just one necessary condition among others. But if one takes this reply, AF becomes unhelpful, as it will not explain the limitations on responsibility that we think exist. Proponents of AF will have to offer further necessary conditions, and thus far, such conditions have not been offered.
outcomes in the appropriate way. Nevertheless, it is a difficult task, and it is not at all clear how this might be done.

The difficulty in specifying the outcome here does not apply only to AF. CHF and SHF must also address how to specify outcomes, and this is a challenge for proponents of these conditions. A challenge is not an argument, however, let alone a knockdown argument. Still, there is a deep worry about requiring actual foresight for derivative responsibility. Agents frequently do not foresee outcomes, either because they are unreflective or simply lack imagination. Suppose Brad did not foresee that he might hit some vehicle, or even the more general outcome that he might harm someone. We do not think that Brad is thereby excused from responsibility. AF seems to favor agents who are unreflective and unimaginative. But this seems the wrong result. If agents are derivatively responsible only for potential outcomes they actually foresee, many of our responsibility ascriptions and practices are unjustified.

4.4 Could Have Foreseen

The fact that AF does not seem to match many of our responsibility ascriptions and practices is a serious mark against it and a strong reason to think that actual foresight is not required for derivative responsibility. Perhaps what is required is not that the agent actually foresee some potential outcome, but rather that the agent could have foreseen some outcome. Even if Brad did not foresee that he might hit a vehicle or harm someone as a result of driving drunk, he could have foreseen this. And perhaps this is what is required for derivative responsibility.

CHF expands the scope of what it can capture, but it does so by distancing the outcome from the agent. AF required a clear connection between the outcome and the agent: the agent is responsible for some outcome because the agent actually foresaw some outcome. The fact that the agent could have foreseen some outcome removes the outcome from the agent’s actual mental life. The fact that the agent could have foreseen some outcome when she acted intentionally, one might object, does not tell us about the agent’s quality of will.

The fact that an agent could have foreseen some outcome is, on its own, rather uninformative and toothless. Just because Brad could have foreseen that he would or might hit another vehicle or injure someone while driving home does not connect Brad with the outcome in a way that explains why Brad is responsible for the outcome that results. If an agent has no
special reason to foresee some outcome, it seems difficult for the volitionist to hold the agent responsible for that outcome. And this seems largely to be a matter of the agent’s quality of will; there is no reason to think the agent has any negative quality of will in simply failing to foresee something that he could have foreseen.\textsuperscript{61}

The volitionist might respond that one might not have a negative quality of will in this case, but one does show a lack of due regard (see Strawson 1993). Nevertheless, one does not seem to show a lack of due regard simply in failing to exercise some capacity. There must be some normative standard that the agent fails to satisfy in their failure to fulfill this capacity. Even if Brad did not foresee that he would or might harm someone, he should have foreseen this. Brad could reasonably have been expected for foresee this outcome. But notice that this is different from appealing to a mere ability of the agent. As an informative epistemic condition, CHF seems inadequate. We would do well to explore a fuller statement of the epistemic condition that incorporates this normative standard. If AF is too strong and CHF is not full enough, perhaps SHF is a better, more useful alternative.\textsuperscript{62}

\textbf{4.5 Should Have Foreseen}

AF was a natural condition for derivative responsibility because of the way it clearly connects some outcome with the agent. In order for SHF to be a plausible contender, it must connect the outcome to the agent as well. If it does offer this connection, it must be found in the “should” that sets it apart from CHF. Some understand this “should” in terms of procedural epistemic obligations one must discharge or certain precautions one must take (Rosen 2004). Rosen writes:

Any given precaution is ‘designed’ to prevent a certain range of harms. That is, when a given precaution is required, it is required in part \textit{because} it tends to reduce the likelihood

\textsuperscript{61} Again, this objection understands CHF, (1), and (2) as jointly sufficient for derivative responsibility. If this isn’t correct, and CHF is merely one necessary condition that should be supplemented with others besides (1) and (2), then one might think CHF need not include some further requirement. This is correct, but the larger point is that CHF is not a particularly useful epistemic condition on derivative responsibility if it must be supplemented in this way.

\textsuperscript{62} As I understand it, CHF is similar to Vargas’s KC:

\textit{(KC)} For an agent to be responsible for some outcome (whether an action or consequence) the outcome must be reasonably foreseeable for that agent at some suitable prior time. (2005: 274)

We can make this sort of condition stronger and more useful by including some normative standard as in SHF, as I explain in 4.5 below.
of a certain vaguely specified class of bad outcomes. In such cases, the harm to be prevented is said to fall within the scope of the risk one takes when one omits the precaution. (2008: 604)

When one undertakes some activity, one has certain epistemic obligations to be informed about the nature of the activity, to understand its risks and its effects, etc. If the agent is ignorant of these risks, or fails to foresee some outcome simply because he failed to take the appropriate steps of reflection and consideration, then the agent is still responsible for this outcome, according to SHF. In this respect, SHF is an improvement over AF.

In DUI, Brad has a duty to understand the effects of alcohol use on his body and the risks associated with this alcohol use. The simple fact that Brad did not foresee that he would or might injure someone when driving drunk doesn’t get him off the hook if he should have foreseen this possibility. If he should have educated himself properly before drinking and driving, then Brad is still responsible for the harm he causes when he drives drunk, according to SHF.

Rosen understands this ignorance of some outcome (or the moral status of some action) as compared to the standard of a reasonable agent. The question at issue in SHF is “whether the agent’s ignorance derives from a failure to do what any reasonably prudent person in his circumstances would [have] done in order to see to it that he was adequately informed” (2004: 301). A similar standard is used in law. But George Sher raises a concern with such a standard, as it is difficult to see how the agent is responsible for some outcome when she has failed to do what a reasonable person would have done (2009: 80 ff.). If the agent is responsible for some unforeseen but reasonably foreseeable outcome, then we want to know how the agent failed. Understanding this standard in light of some ideal reasonable agent seems to distance the outcome from the agent in a way that is unacceptable.

A better proposal, then, is to understand these obligations and this lack of foresight not as relative to some reasonably prudent person, but instead as relative to the agent and her capacities. In other words, the “should” in SHF should be read as relative to the individual agent and her capacities, such that it was reasonable to expect the agent with those capacities in that situation to foresee that some outcome would or might occur. Thus, when we say that Brad should have foreseen that he would or might harm some people as a result of driving drunk, we mean that, given various facts about Brad’s actual psychology, including his mental capacities
and moral imagination, as well as the situation he was in, it is reasonable to expect him to foresee this outcome.

SHF casts the net of derivative responsibility much wider than AF, and it connects the outcome to the agent better than CHF by appealing to a normative standard that is relative to the agent’s capacities. SHF seems poised to explain an agent’s responsibility for a variety of items that VT2 does not capture, including attitudes and character traits. In fact, some writers have appealed to something like the tracing method when explaining how agents can be responsible for their emotions (Sankowski 1977). Supposing that an agent performed some intentional action and the agent foresaw or should have foreseen that some emotion would or might result, the agent could be derivatively responsible for this emotion.

The same might be said of character traits. If Dan, from *Fellowship*, chose to engage in certain behaviors when he was younger, and he foresaw or should have foreseen that these behaviors would lead to him being resentful of others who are more successful than him, then Dan is derivatively responsible for bringing about his resentment. Perhaps Dan chose to surround himself with competitive and jealous people. He may have performed some mental action that resulted in an important omission, such as a choice not to meditate or reflect on his own successes. If we can pinpoint a previous intentional action for which Dan was directly responsible and at which time he foresaw or should have foreseen that this action would bring about some outcome of a type like that in *Fellowship*, then Dan is responsible for his scowl when John wins the fellowship. Of course, the volitionist also holds that if there is no such previous intentional action of Dan’s such that SHF (or whichever condition above is most plausible) is satisfied, then Dan is not responsible, either directly or derivatively, for scowling when John is announced as the winner.

Clearly, whether the volitionist will attribute responsibility to an agent for some item will depend on the details of the case. Unfortunately, most cases are woefully underdescribed. Recall *Sore Tooth*, in which the politically liberal Fiona feels a pang of dismay when she realizes that her new dentist is an African-American. The volitionist holds that Fiona is not directly responsible for her racist attitude. In order to determine whether Fiona is derivatively responsible

---

63 As I discussed in Chapter 1, some volitionists may deny that we are even derivatively responsible for these items. While it may be the case that we are not responsible at all for our attitudes or character traits, this conclusion would require a drastic revision of our current responsibility ascriptions and practices. I assume that a volitionist would prefer to hold agents responsible for these items if possible in order to avoid such a revisionist view.
for her racist attitude, we must know whether there is some intentional action, $A$, that Fiona performed such that she foresaw or should have foreseen that $A$ would or might contribute to the outcome that she would feel dismay at the color of someone’s skin. But *Sore Tooth* gives no such details. We could tell a story such that Fiona chose to surround herself with racist friends, foreseeing that spending time with such people might have negative effects on her own character and the attitudes she holds. Perhaps despite opportunities to spend time with people of other races and cultures, she chose not to take such opportunities. Given these details, the volitionist might claim that Fiona is derivatively responsible for her racist attitude, despite her claims that the color of a person’s skin does not affect a person’s worth. This is because Fiona’s actions in the past serve as an anchor from which we can derive her responsibility for her racist attitude.

Now imagine that the details of *Sore Tooth* are different. Fiona has not made any choices to spend time with racist people, nor has she chosen to pass up opportunities to learn about people of other races or cultures. In fact, Fiona has intentionally taken psychology courses to learn about the implicit social biases she has from growing up in the United States. She has worked to correct these biases as best she can. In this case, there is no previous intentional action of Fiona’s to which we might trace responsibility for her racist attitude. As a result, the volitionist holds that Fiona is not even derivatively responsible for her racist attitude. It is unfortunate that she has this objectionable attitude, and perhaps it is bad that Fiona has the attitude. Nevertheless, says the volitionist, Fiona is neither responsible nor blameworthy for having the attitude.

It would appear that SHF is poised to serve as a nice corollary to VT2 such that the volitionist can account for our everyday responsibility ascriptions and practices either directly or derivatively. Yet there are two difficulties with SHF. The first is familiar: SHF requires that its advocates specify the outcomes in SHF such that they are neither too narrow nor too broad. While specifying the outcomes appropriately in AF is difficult, specifying those outcomes in SHF is even more difficult. While an agent might not actually foresee a broadly specified outcome, such as “I will treat some people poorly in the future as a result of this intentional action,” it might be reasonable to expect the agent to foresee such broadly specified outcomes. That is, even if the agent does not foresee some broadly specified outcome, it is far more likely

---

64 I am not imagining that Fiona chose to not take such opportunities because she thought that she was of a superior race, but rather for some unrelated reason that does not constitute an excuse.
that the agent *should* have foreseen some broadly specified outcome. Thus the problem of too much responsibility looms even larger for the proponent of SHF.

The second difficulty concerns the procedural epistemic obligations one must discharge, which is the basis of the “should” in SHF. It seems possible that agents may frequently fail to fulfill these epistemic obligations, and these failings might not be intentional actions or traceable to any previous intentional action. Nevertheless, we do tend to hold agents responsible in such cases, thinking they deserve blame if the failure is negative.

Compare two cases:

*Airport 1:* Jack promised to pick up Kelly from the airport at 8:00. At 7:00 he is engrossed in a television show, so he deliberates about whether to go to the airport or stay and watch the show. He decides not to go to the airport, and 8:00 comes and goes without Jack leaving the house.

*Airport 2:* Josh promised to pick up Kate from the airport at 8:00. Despite his promise and his generally reliable memory, he forgets about picking up Kate, and 8:00 comes and goes without Josh leaving the house.

In both cases, I suggest, we are inclined to blame the person who fails to fulfill his promise to pick up someone from the airport. I also think that we believe such blame to be justified. In *Airport 1,* the justification for this blame is rather straightforward. According to VT2, Jack is not directly responsible for his omission because this omission is not an intentional action. It isn’t an action at all. Nevertheless, Jack’s decision not to go to the airport but to stay and watch the television show is an intentional action, and at the time of this intentional action he foresaw (or should have foreseen) that by staying and watching the show, he would fail to fulfill his promise to pick up Kelly from the airport. Jack may be derivatively responsible for his omission in *Airport 1* because he satisfies SHF.

The story is more complicated in *Airport 2.* Josh cannot be directly responsible for his omission, as it is not an intentional action. If Josh is responsible, it must be in virtue of some prior intentional action he performed during which he foresaw or should have foreseen that he would or might fail to pick up Kate. But in the case, Josh made no prior decision not to pick up
Kate. What intentional action did Josh perform during which he foresaw or should have foreseen his failure to pick up Kate? Where might we anchor Josh’s responsibility?

One might suggest that it is Josh’s epistemic duty to leave himself a note as a reminder to pick up Kate. In failing to leave himself a note, Josh failed to discharge his epistemic duty, and it is in virtue of this failure that Josh is responsible for not picking up Kate. It is not clear, however, that Josh is required to leave a note given his generally reliable memory. Requiring that Josh leave a note seems excessive as an epistemic obligation.

In any case, suppose that Josh made no decision not to leave a note. He simply forgot to leave a note for himself reminding him of his promise to Kate. Should Josh have left a reminder to himself to leave a reminder? Again, this seems excessive. How far back should we trace Josh’s responsibility, and what might possibly ground it? There seems to be no intentional action Josh performed during which he foresaw or should have foreseen that he would or might fail to pick up Kate. Airport 2 is an apparent nontracing case—a case in which the agent is responsible for some item and yet responsibility for that item cannot be traced back to some prior intentional action.65 If Josh is responsible for failing to keep his promise to Kate and yet there is no way to trace this responsibility to some previous intentional action in a way that satisfies AF or SHF, it seems Airport 2 is a counterexample to both necessary conditions on derivative responsibility.

A volitionist is free to claim that Josh simply isn’t responsible in Airport 2. This bears some cost: we do, as a matter of practice, tend to hold people responsible in such cases. To the extent that there are a great number of such cases, volitionists must bear a substantial cost to their theory. And it seems such cases can be multiplied. In addition to simple forgetting cases, volitionists might have a difficult time explaining why agents are responsible for some omission when that omission is not preceded by some intentional action, such as a decision. Perhaps Mirja stands on the sidewalk as a friend picks up spilled paperwork. Mirja does not choose not to help, nor does she choose to stand still. She simply makes no choice at all. She takes no action. It seems difficult to see how Mirja could be responsible for her omission either directly or derivatively according to the proponent of VT2, as the omission is not an action, nor is it the reasonably foreseeable outcome of any previous intentional action. What previous intentional action might Mirja have performed at the time of which she should have foreseen that she might

65 Airport 2 is similar to Birthday in this regard. We might suppose that Angela forgot to leave a note to call Dave or that she did not perform some previous intentional action that would ground her responsibility for her omission.
fail to help someone in this way?\textsuperscript{66} It is an important challenge to volitionists that just as many of our daily actions are not the result of prior deliberation or choice, many of our omissions likely fall into this category as well. Given how often we omit to act without making such choices, the volitionist has cause to be uneasy about this failure of VT2 and SHF.

Some writers have argued that cases of negligence pose a special problem for tracing in general. Consider the following case, taken from Matt King:

\textit{Negligence:} Nate, tired from waking up early, is backing out of the driveway. His thoughts turn to his meetings that day, and his attention is partially focused on a radio commercial. Due to his inattention, Nate doesn’t see a child walking to school and so hits him, breaking the child’s leg. (2009: 578)

Nate was negligent in that he failed to pay proper attention to what he was doing. In so failing, Nate risked harm to others. King notes that we seem to think that Nate is responsible and blameworthy for his negligent conduct. He did not pay attention, and consequently he is responsible for the child’s broken leg.

Nate’s hitting the child was not an intentional action. The outcome that the child’s leg is broken is also not an intentional action. Nate is not directly responsible for hitting the child or for the outcome that the child’s leg is broken. If Nate is responsible, then, the volitionist must hold that he is derivatively responsible. But to where should we trace Nate’s responsibility for hitting the child? As King explains, Nate did not choose to be negligent. He also does not seem to be responsible for creating his general condition of inattentiveness. We might assume that the thoughts of his meeting simply spring to mind, as many of our thoughts so often do. We also need not assume that Nate chooses to be groggy, as he may have gone to bed at a decent hour the night before but nevertheless be tired. Nate’s choice to turn on the radio is not the same as a choice to be distracted by the radio. He may simply find that a commercial grabs his attention in

\textsuperscript{66} Perhaps one might insist that Mirja might have made some previous choice or performed some previous intentional action during which she foresaw or should have foreseen that she might fail to help someone. It is not clear to me what this intentional action might be. Mirja might have decided to make herself more selfish or independent, and in performing this action she might have foreseen (or should have foreseen) that she would fail to help her friend. Yet this seems a rather unlikely intentional action for Mirja to perform. In addition, the specification of outcomes once again becomes important. “Failing to help someone” is quite broad, and it seems that on this specification of the outcome, Mirja might be responsible for too much. Yet clearly Mirja did not foresee nor should she have foreseen that she would fail to help a friend pick up papers on a sidewalk.
much the same way certain thoughts rise to our conscious attention. There seems to be no good place for the volitionist to trace Nate’s responsibility. Nevertheless, Nate seems to be responsible for hitting the child and for the fact that the child’s leg is broken.

These sorts of cases suggest that the proponent of VT2 will have a difficult time accounting for an agent’s responsibility for many items for which we hold agents responsible, including cases of forgetting, negligence, and certain types of omissions. Agents cannot be directly responsible in such cases, but it also seems that they cannot be derivatively responsible. A volitionist can either offer a different way to account for responsibility in these sorts of cases or else accept that agents simply aren’t responsible in such cases. It is difficult to see how volitionists might accomplish the former. While the latter is consistent with their theory, one might reasonably wonder whether there is some other justification for claiming that agents like Josh and Mirja are responsible.

4.6 Taking Stock

The proponent of VT2 must account for our responsibility for items that are not intentional actions. The best way to do this is derivatively, appealing to either AF, CHF, or SHF. AF seems too narrow, and CHF seems an uninformative and toothless condition. SHF does seem to be an improvement. Nevertheless, SHF fails to capture all of the items we typically hold agents responsible for, and it seems especially ill-equipped to explain responsibility in cases of forgotten promises, negligence, or omissions that were not proceeded by some choice. Whether the volitionist opts for AF or SHF, there remain important difficulties in specifying how outcomes ought to be understood in either condition. Should outcomes be understood too narrowly, agents will not be responsible for items for which they do seem to be responsible. Yet should outcomes be understood too broadly, agents will be responsible for items for which they do not seem to be responsible. As of yet, proponents of tracing have not offered a way to specify outcomes such that all and only the outcomes for which we hold agents responsible are captured.

None of these concerns are knockdown objections to VT2, AF, or SHF. Instead, they are important challenges that a proponent of VT2 must address and overcome. Importantly, these challenges do not arise simply as a result of understanding the volitionist thesis as VT2. Often, the intentional action that serves as the anchor of responsibility is the decision or choice the
agent makes, which coincides with prior choice and endorsement views. Rejecting VT2 as the thesis that best represents these families of volitionist views does not solve the problem.

This is especially noteworthy because some theorists, like Fischer and Tognazzini, see themselves as being immune to certain attacks on tracing cases because these attacks focus not on choice, which they call a “volitional approach” to responsibility, but on control: “Whereas a volitional approach requires that we trace back (directly or indirectly) to a choice in order to justify attributions of moral responsibility, a control model requires that we [trace] back (directly or indirectly) to an exercise of control, where such an exercise of control might be a choice, and [sic] action, or an omission” (2009: 549). I have argued in chapter 2 that this volitional approach, as Fischer and Tognazzini call it, does not seem to be what volitionists have in mind—or at least, it shouldn’t be what volitionists have in mind. VT2 is much broader than this volitional approach because it appeals to intentional actions rather than just a subset of mental actions like choice. So it seems that the so-called volitional approach is simply a straw man.

If the notion of control Fischer and Tognazzini have in mind for their control approach is guidance control, the considerations I raised in chapter 3 suggest that this disjunction of a choice, an action, or an omission might not involve an exercise of the same sort of control. If not, then appealing to a control approach might not be as clear-cut as they hope. They claim, “[I]f we replace volition with a broader notion of control, where the pertinent sort of control can be exercised via an omission as well as (say) a choice, one can give an adequate and, indeed, a natural account of [cases like Birthday]” (551). Unfortunately, there does not seem to be some unified “pertinent sort of control” exercised in both actions and omissions, and even if there were, the problems I raised above for forgetting cases and cases of negligence still seem to apply. Fischer and Tognazzini’s attempt to resist such cases (in this way, at least) is unsuccessful.

Although the above considerations are not, individually, knockdown objections to VT2, AF, or SHF, together they pose a significant challenge. In addition, there is a further reason to be skeptical of the power of tracing cases that arises from our practices of holding agents responsible. Angela Smith argues that even if we can account for various attitudes, character traits, and other outcomes derivatively, this is often not how we in fact hold agents responsible:

While it may be true that we could, over time, exert an influence on many of these patterns of awareness and response by voluntarily pursuing various projects of self-
improvement, I do not think that this fact plays a role in explaining either our willingness to take these attitudes as a legitimate basis for (moral) appraisal or the nature of the appraisals we in fact make. When we praise or criticize someone for an attitude it seems we are responding to something about the content of that attitude and not to facts about its origin in a person’s prior voluntary choices, or to facts about its susceptibility to influence through a person’s future voluntary choices. More specifically, it seems we are responding to certain judgments of the person which we take to be implicit in that attitude, judgments for which we consider her to be directly morally answerable. If this is correct, then it is a mistake to try to account for a person’s responsibility for her own attitudes in terms of their connection to her prior or future voluntary choices, because that obscures the special nature of our relation to our own attitudes: we are not merely producers of our attitudes, or even guardians over them; we are, first and foremost, inhabitants of them. They are a direct reflection of what we judge to be of value, importance, or significance. (2005: 251)

AF or SHF may tell a plausible story of how an agent is derivatively responsible for having some attitude, but when we blame an agent for her morally objectionable attitude, Smith claims, we are not responding to the attitude’s origin. We are responding to that attitude and what it expresses about the person, holding that person to be directly responsible for the attitude. A morally objectionable attitude reveals something about the agent herself, according to Smith, and the volitionist technique of tracing the agent’s responsibility to some prior intentional action ignores the direct connection between the agent and her attitude.

In part, Smith’s objection requires some empirical verification. We would like to know whether we do tend to hold agents responsible via some sort of tracing technique. It seems plausible that at least sometimes we do. Consider our typical reaction to a racist attitude in some agent, Meryl. We blame Meryl, thinking that her racist attitude reflects poorly on her. But imagine that we learn that Meryl was raised in the American South by grandparents with deep-seated racist views. Despite her racist upbringing, which we can stipulate was not indoctrination or brainwashing, Meryl has undertaken the process of trying to remove many of her racist attitudes. She is mindful of her biases and engages in conscientious techniques designed to combat these biases and remove these racist attitudes. While she still has some racist attitudes,
on the whole she has far fewer racist attitudes than she used to. Upon learning this information about Meryl, we might very well claim that Meryl is not blameworthy for her racist attitude, or not very blameworthy for her racist attitude. We trace Meryl’s responsibility for her racist attitudes back to some previous intentional action, and if we can find no intentional action to which we can anchor responsibility, our inclination to hold Meryl responsible fades.

Compare Meryl to Daryl, who intentionally immersed himself in racist literature, foreseeing that this immersion would or might cause him to have racist attitudes. Unlike Meryl, we are not at all reluctant to hold Daryl responsible for his racist attitudes, but we do so derivatively.

While Meryl and Daryl suggest that Smith is mistaken, I think there is more to this story. Sometimes we do seem to hold agents responsible for attitudes derivatively, in virtue of previous choices they have made. This seems to be the case for Meryl and Daryl. Yet we do not always hold agents responsible for their attitudes or certain beliefs derivatively, nor do we seem to think such a judgment is necessary. I, for one, blame Donald Trump for his objectionable attitudes not in virtue of previous choices that he made to have those attitudes. Rather, it seems that Trump is responsible for these attitudes more directly in virtue of what these attitudes reveal about him as an agent. These attitudes seem to reflect Trump’s values—values that are offensive and unjustifiable.

Similarly, we do not blame someone for not writing a reminder when he fails to pick us up from the airport. We blame him more directly, simply for the failure. Notice that this is different from the way in which we blame a drunk driver. We do blame the drunk driver for drinking to excess, foreseeing that he will have to drive home and may harm someone. This previous event seems to be the locus of responsibility in the drunk driving case. Yet there is a disconnect in cases of negligence; we do not search for some previous instance at which an agent should have left a note. Instead, we blame the agent simply for the negligence and for what this negligence reveals about the agent.

If these considerations are correct, it seems that there are three important reasons to be skeptical of the power of tracing and of volitionism’s ability to explain our responsibility for many attitudes, character traits, and omissions—at least in certain cases. First, on the assumption

---

67 Given what I argue in chapter 6, this also seems an appropriate result. I do not think that agents can be directly responsible for implicit racist attitudes, and it seems that Meryl’s attitudes are implicit.
that direct responsibility is limited to intentional actions, neither AF nor SHF seem able to explain our responsibility for all such cases. Second, specifying the outcomes in such a way that all and only the right outcomes are captured by AF or SHF is either a difficult task that has not yet been completed or an impossible task that cannot be completed. Finally, even if tracing is successful in some cases, as a matter of practice we tend to hold agents responsible for attitudes and negligence in a more direct way. Given these reasons, we should be motivated to at least entertain other accounts of responsibility that are not volitionist accounts. Perhaps such nonvolitionist accounts can fare better than these volitionist accounts in matching our everyday responsibility practices and ascriptions.
CHAPTER 5

A NONVOLITIONIST ALTERNATIVE:
THE RATIONAL RELATIONS VIEW

5.1 Introduction

I have argued that volitionists who embrace VT2 cannot hold that agents are directly responsible for many intentional omissions. VT2 concerns only intentional actions, and many omissions are not actions at all. While volitionists might be able to argue that we are derivatively responsible for some intentional omissions (among other items), this tracing technique faces several challenges and seems to fail in many cases of negligence or failures to notice certain features. As a result, our responsibility for many omissions cannot be captured at all on this volitionist picture. This is an unhappy result, as we do, in practice, hold agents responsible for many omissions, including negligence and failures.

A volitionist might attempt to modify VT2 to VT4, but this seems unjustified. While adopting VT3 is tempting, it seems that quality of will is important in explaining responsibility. VT3 on its own does not include any such appeal. Consequently, volitionists are left with a choice. They can adopt VT2 and accept that we are responsible for much less than we ordinarily think. Alternatively, they can adopt VT3 to capture direct responsibility for more items. If they adopt VT3 alone, they sacrifice a plausible grounding for this responsibility. If they supplement VT3 with a quality of will condition, they bring their view much closer to nonvolitionism.

Nonvolitionists have been eager to argue that volitionists cannot adequately account for responsibility for nonvoluntary items such as attitudes. They claim that a view that cannot accommodate these items is unacceptably revisionist, and this desire to avoid a revisionist picture of responsibility seems to motivate most nonvolitionist theories of responsibility. If my arguments from chapters 3 and 4 are on track, however, volitionists plausibly struggle to account for responsibility not only for these items, but also for many intentional omissions as well as many standard cases of negligence, ignorance, and unwitting omissions for which we typically believe agents are responsible. This is a point that volitionists have not fully appreciated, and
nonvolitionists can bolster the motivation for their view by appealing to these arguments. To the extent that a theory includes too many items we want to rule out or excludes too many items we want to rule in, the theory has some revisionist cost. Given that we typically do hold agents responsible for many attitudes, intentional omissions, negligent acts, and even unwitting omissions, if there is a theory that can match our ordinary practices and ascriptions and adequately explain them, we should prefer this theory to a more revisionist theory.

In light of the shortcomings of volitionism, then, nonvolitionism might seem more plausible. But we should not endorse a theory unless it can adequately explain how an agent is responsible for these nonvoluntary items. As we have seen, VT3 alone struggles to explain what grounds an agent’s responsibility. Can nonvolitionists offer a better account of how and why agents are responsible for many attitudes, beliefs, intentional omissions, and unwitting omissions?

In this chapter, I survey and defend one of the most well-developed and notable nonvolitionist views: Angela Smith’s rational relations view (RRV). In the process, I also examine George Sher’s responsibility without awareness view (RWA). RWA is a wider view that aims to explain how agents can be directly responsible for acts or omissions even when they lacked awareness of such acts and omissions. I argue that RRV is preferable over RWA, because much like unsupplemented capacity views, RWA offers little in the way of grounding responsibility. RRV can meet this challenge, however.

5.2 The Rational Relations View

If some connection to choice or intention cannot ground our responsibility for all the items for which we typically hold agents responsible, what might ground these items? According to Angela Smith, the answer to this question lies in the expression of our moral agency: “[W]e express our moral agency and activity not only in our explicit choices and decisions, but also in what we unreflectively think, feel, desire, and notice. Theories which make choice or voluntary control a precondition of moral responsibility, in my view, leave us with an impoverished

---

68 I will refer collectively to attitudes, unwitting or unintentional omissions, negligent behaviors, and failures to notice or remember as nonvoluntary items (or nonvoluntary faults when they are bad or objectionable). This is intended as a term of art for ease of reference, though I appreciate the difficulty in specifying what counts as voluntary and nonvoluntary.

69 Sher’s view seems to straddle the line between volitionist views and nonvolitionist views, so I avoid the use of either label in this case. For more on this problem, see note 89 below.
conception of moral personhood” (2005: 263). We have already seen that volitionist views often do not capture responsibility for all the items we might want them to. According to Smith, the problem is that volitionists focus on only one way in which our moral agency is expressed. But, she claims, our evaluative moral judgments are also expressed in various nonvoluntary items, including our attitudes.

Smith is one of a group of nonvolitionists typically called attributionists. Attributionists claim that an agent’s intentional actions are not the only ground of her responsibility. Instead, the fact that some item expresses an agent’s evaluative judgment is sufficient to explain why that agent is responsible for that item. Proponents of attributionism include T. M. Scanlon 1998 and 2002, Angela Smith 2005 and 2008, Matthew Talbert 2008 and 2012b, and Pamela Hieronymi 2008 and 2014. Because Smith’s view is arguably the most influential and most developed, it will be my focus here. Much of what I say should apply equally well to other varieties of attributionism, however.

Smith calls her flavor of nonvolitionism the rational relations view (RRV). According to Smith, “To say that an agent is morally responsible for some thing, on this view, is to say that that thing reflects her rational judgment in a way that makes it appropriate, in principle, to ask her to defend or justify it” (2008: 369). To understand RRV, we must be clear about what a rational judgment is, how it is reflected in some item, and how a defense or justification is, in principle, appropriate.

Let’s begin with the judgments at issue. Sometimes Smith refers to these judgments as rational judgments, while at others she refers to them as evaluative judgments. But these terms seem to be interchangeable on her view. An evaluative judgment, for Smith, is not necessarily a consciously held propositional belief or explicit judgment, but rather a tendency “to regard certain things as having evaluative significance” (2005: 251).

Perhaps better is the idea that intentional actions are not the fundamental grounding of an agent’s responsibility, as these intentional actions can be further grounded in the agent’s evaluative moral judgments.

There are important differences between the views of these writers, of course, and not all of these writers might self-identify as attributionists. (In fact, Smith argues that the term “attributionist” is misleading and problematic in important ways.) Nevertheless, exploring the differences between these views and whether “attributionist” is an appropriate term for some view are both projects tangential to my purpose here.

Perhaps Smith’s sliding between these terms is due in part to the influence of Scanlon 1998 (chs. 1 and 6). Smith acknowledges that her focus on judgment-dependence and the appropriateness of demanding reasons or justification come from Scanlon (Smith 2005: 253, n. 30). From here on I will generally use “evaluative judgment” over “rational judgment” for consistency.
Many nonvoluntary items, including attitudes as well as physical states and behaviors, are rationally connected to these evaluative judgments. This rational connection reflects a normative ideal:

[S]ome of our mental states are linked to particular judgments in such a way that, if one sincerely holds a particular evaluative judgment, then the mental state in question should (or should not) occur. The ‘should’ in question here is the should of rationality and, therefore, marks a normative ideal which our actual attitudes may not always meet. (2005: 253)

It is at least partially in virtue of this rational connection that an agent’s attitudes, for example, reflect or express her evaluative judgments. To take a fairly straightforward example, one might see a man standing in the doorway with a gun and judge that the man and weapon are dangerous. This judgment should, rationally, give rise to certain attitudes—most notably, fear.

Not only should certain attitudes occur if we hold some particular judgment, but if we have some certain attitude, then there should be some underlying judgment that explains this attitude. We can see this in a variety of attitudes, including contempt, guilt, resentment, and admiration. Resentment reflects the judgment that one has been wronged. One’s desire to perform some action reflects the judgment that that action is good in some way. If one did not hold this judgment that one has been wronged, one would not have the attitude of resentment. Given that one does have this attitude, one must, in fact, hold this judgment. In virtue of the normative connection between attitudes and judgments, Smith says, “we can make a direct inference from the occurrence of such states to the underlying judgments these responses reflect” (254).

Consider an agent who has an attitude of disdain toward Muslims. This attitude might reflect a judgment that Muslims are intellectually inferior because of their religion. This judgment in turn may reflect the morally objectionable belief that one’s intelligence is determined by religion and that one’s moral worth is based on her intelligence. But we can, according to Smith, make this inference to the agent’s underlying evaluative judgments given the

---

attitudes the agent has. A variety of nonvoluntary items, including many attitudes and behaviors, are rationally connected to our evaluative judgments in this way.

Similarly, says Smith, we can make inferences about a person’s evaluative judgments given a lack of attitudes:

[I]f a person judges something to be important or significant, this should (rationally) have an influence on her unreflective patterns of thought and feeling concerning the thing in question. … [I]f a person fails to notice or respond in the relevant ways, we take this as at least some indication that she does not accept the judgment in question. (2005: 255)

When Angela forgets to call Dave in Birthday, we can take her behavior as an indication that she does not judge her promise or Dave’s birthday to be particularly important. Often, such noticing or neglecting and occurring to us are similarly connected to underlying evaluative judgments.

Already we can see that a variety of nonvoluntary items as well as voluntary items are rationally connected to our evaluative judgments. According to Smith, the reason why it is appropriate to hold an agent morally responsible for these items in the sense in which we are interested is because “moral criticism addresses a person qua rational agent and asks her to acknowledge and to defend or disavow the judgments implicit in her responses to the world around her” (256). For Smith, this connection between evaluative judgments and the items to which they give rise is precisely the kind of connection that justifies the sort of reactive attitudes we take to be connected to moral responsibility.

Of course, on RRV, not every nonvoluntary item is one for which an agent can be held responsible. Smith writes, “In order for a creature to be responsible for an attitude, on RRV, it must be the kind of state that is open, in principle, to revision or modification through that creature’s own processes of rational reflection” (256). The focus here is on a rational connection; if some attitude or state is not answerable to a person’s judgment via this rational connection, then the person isn’t responsible for that item.

To summarize, according to RRV, some of our mental states—including many of our intentions, beliefs, and other attitudes—reflect our evaluative judgments. There is a rational connection between these judgments and the items to which they give rise (or fail to give rise). Given that there are these normative connections between the judgments and these items, we can
justify these mental states by giving reasons, and we can legitimately ask agents for such reasons. Being responsible involves being a target of these normative demands, and since it is appropriate to demand reasons for these items in virtue of their rational connection to evaluative judgments, agents are directly responsible for these items (270-271).

RRV is clearly different from volitionism not only in the scope of that for which agents can be directly responsible, but also in the way in which agents are responsible for those items. As Smith understands the difference, “Volitionists claim that a person must have voluntarily chosen to bring about an action or attitude in order for her to be responsible for it; the rational relations view claims that a person’s judgment must be reflected in an action or attitude in order for her to be responsible for it” (2008: 371). In other words, for Smith, responsibility isn’t explained in virtue of a connection to voluntary control or intentional action. Instead, one could say that responsibility is explained in virtue of a connection to rational control. RRV seems to capture much more than VT2 and VT4, and it does so in a way that requires no ad hoc justifications. In addition, RRV appeals to the normative connections between an agent’s mental states and behaviors and an agent’s evaluative judgments, so it provides more grounding than VT3 alone.

5.3 Objections to the Rational Relations View

Unlike volitionist views, RRV seems able to explain how agents are directly responsible for many nonvoluntary items. As a result, it has drawn a lot of attention. Although some theorists, like Matthew Talbert 2013, have essentially adopted or slightly adapted RRV, others have found it problematic for a variety of reasons. In this section, I examine some common objections to RRV, along with what I take to be plausible responses. While I think that none of these objections poses a devastating challenge to RRV, exploring these objections and some responses can help us to better understand what RRV entails.

5.3.1 Too Close to Volitionism

Some theorists, such as Neil Levy 2005 and Holly Smith 2011, have criticized RRV for being too close to volitionism, or for implicitly appealing to volitionist requirements. According to this objection, although Angela Smith denies that agents need the control with which volitionists are concerned, she ultimately ushers in this control through the back door. Both Levy
and Smith are focused on the following passage from Angela Smith: “In order for a creature to be responsible for an attitude, on the rational relations view, it must be the kind of state that is open, *in principle*, to revision or modification through that creature’s own processes of rational reflection” (2005: 256, emphasis mine). Here, it sounds as though Angela Smith is claiming that an agent must, *in principle*, be able to control her attitude in some way in order to revise that attitude. This reading has led Holly Smith to believe that “this restriction brings her uncomfortably close to insisting on the very kind of voluntarism that she rejects among her opponents” (2011: 118, n. 9).

Levy pushes the point more forcefully, noting that the appeal to “*in principle*” control is even less helpful than a straightforward appeal to an exercise of control or a capacity for control:

But the attributionist claim is that I am responsible for my judgment-sensitive attitudes not because I have—or even can—control them, but because they are *in principle* ‘under the control of reason’ (Scanlon 1998: 272). It need not be the case that I have controlled them, or even that I could control them, for me to be responsible for them (276). I can see good reasons for thinking that actual control matters morally. But I see none for thinking that *in principle* control matters at all. I don’t have a kind of ersatz control over my car if the steering wheel falls off; the fact that cars are *in principle* controllable does not alter my lack of control in that particular circumstance. If control matters, then its absence cancels responsibility (unless of course the agent is responsible *for* her absence of control). If it does not matter, then we ought not to appeal to it. Appealing to *in principle* control is appealing to an unstable and arbitrary halfway house. (2005: 10)

Levy’s thought is that control is required for responsibility on any plausible account. Although nonvolitionists like Scanlon and Angela Smith seem to reject this requirement, Levy thinks, they still apparently recognize the importance of control in responsibility, as they appeal to *in principle* control. An attitude may be, in principle, under the control of reason, even if some individual cannot ever control or alter that attitude. But, Levy contends, this in principle control is not relevant to responsibility; responsibility requires something stronger. It requires not in principle control, but an exercise of control or a capacity for control. Nonvolitionists cannot plausibly reject the requirement for an exercise of control or a capacity for control. The fact that
they appeal to in principle control suggests that they recognize the importance of control for responsibility, but they are unwilling to commit fully to this control requirement. The result, Levy claims, is an appeal to a watered-down control requirement that can’t justify responsibility.

The objection that Angela Smith and other nonvolitionists are too close to volitionism because they appeal to control is rooted in a misunderstanding. Smith is not illegitimately implicitly appealing to a notion of control that she has explicitly rejected in virtue of rejecting volitionism. When Levy mentions control, he does not specify what kind of control is at issue. Control is relevant for Smith, but she is not concerned with voluntary control, but rather with rational control. These distinct notions of control explain the differences between RRV and volitionism. As Smith is at pains to demonstrate, volitionists think that our practical agency is expressed only through our intentional actions. Voluntary control is required for responsibility, they claim, because voluntary control is the vehicle by which our practical agency is translated into behavior. But Smith thinks that our practical agency is expressed in our attitudes, rational judgments, and nonintentional or unintentional responses as well. Volitionists rule out these items precisely because they are not intentional actions that are under one’s voluntary control. While Smith is appealing to control, it is not the same kind of control as the volitionist.

What exactly is this rational control, and how is it different from voluntary control? As I understand Smith’s conception of rational control, to say that one has rational control over some item is just to say that the item is open, in principle, to revision via rational reflection. Both attitudes and evaluative judgments can be open to revision via rational reflection. According to Smith, for an attitude, $A$, of some agent $S$ (that reflects one of $S$’s evaluative judgments, $V$) to be open, in principle, to rational revision, it must be the case that “if $S$ were to cease to have $V$, either $A$ would be altered or else $S$ would cease to have $A$” (Miller 2014: 471). For an evaluative judgment, $V$, to be open, in principle, to rational revision, it must be the case that “If $S$ judged that some reason $R$ was a sufficient reason to give up $V$, and if $S$ did not hold any other judgments (conscious or subconscious) that conflicted with $R$, $S$ would cease to have $V$ because of $R$” (472, n. 9). In other words, these attitudes and judgments are normatively connected such that if one no longer holds certain judgments, one’s attitudes or other judgments are revised or affected.

---

74 This clarification comes from Smith herself in personal correspondence with Miller. Miller 2014 argues that RRV needs an historical condition to avoid problematic manipulation cases. Miller offers such a condition, which I think Smith would be wise to adopt. I myself advocate its adoption in the nonvolitionist view I offer in chapter 7.
Consider an attitude like shame, which is constituted in part by some underlying judgment that one has done something wrong. This attitude of shame is open to rational revision, as it is true that if one’s judgment changes, then necessarily the attitude will also change. If one ceases to judge that one has done something wrong, then one no longer can feel shame. Perhaps there is some sort of affect that lingers, but this cannot be called shame without the underlying judgment.

Not all attitudes are constituted by underlying judgments, however. Some are merely caused by such judgments. In these sorts of cases, the connection between the judgment and the attitude is contingent. Suppose that one ceases to hold some evaluative judgment, $V$, and yet one still has some attitude, $A$, that one rationally should not have if one does not hold $V$. $A$ is contrary to the agent’s settled judgment. Is $A$ open, in principle, to rational revision in this case? If one reads Smith’s conditional at face value here, it appears that $A$ is not open to rational revision. The reason, it seems, is that Smith is trying to capture an individual capacity to revise some attitude. One apparently lacks the capacity in this case. If one lacks this capacity, then the attitude is not revisable for the agent, and so lies outside the agent’s rational control. On this reading, Levy’s objection does not apply to Smith’s RRV. Rational control requires that the individual have the capacity to revise her attitudes, though this capacity is a rational capacity, not a volitional capacity.

Smith’s usage of “in principle” is somewhat curious, however, and suggests she might not require a capacity after all. If Smith’s counterfactuals are not to be taken at face value, and she does not require a capacity, then Levy’s objection might find its footing. Suppose the normative connection Smith requires between an attitude and judgment is simply that if one holds $V$, one must have $A$. That connection might still exist even if the agent is incapable of responding rationally to $V$. On this reading, Smith seems quite close to Scanlon, as Levy describes him. Scanlon does not require some sort of capacity, as it need not be the case that one even could control one’s attitudes in order to be responsible for them (Scanlon 1998: 276). If not even a capacity is required, Levy’s objection is more biting, even though he does not distinguish between rational control and voluntary control. The fact that this normative connection exists, even if it is true that my attitude would not alter in response to a change in my judgment, is a poor grounding for responsibility.
Which reading should be understood as Smith’s view is unclear. I find the appeal to a capacity the most promising response. If one requires the capacity, as it seems like Smith does in offering the counterfactuals above, then Levy’s objection that attributionists appeal to an “unstable and arbitrary halfway house” misses the mark. Such nonvolitionists can happily appeal to a capacity for control. This control is not voluntary control or volitional control, but rational control. As a result, these nonvolitionists can distance themselves from volitionism without implicitly appealing to volitionist requirements.

5.3.2 Too Weak to Justify Robust Moral Responsibility

A volitionist might accept that RRV makes no implicit appeal to voluntary control and is adequately distinct from volitionism, but nevertheless reject RRV because rational control simply isn’t robust enough to ground responsibility—at least, not the sort of responsibility that warrants the reactive attitudes (see Levy 2005). Two things are important to note about this objection. First, it is a separate objection from the concern that RRV is a poorly disguised volitionist view, and it seems to require that one accept that RRV and volitionism are substantially different and rely on different notions of control. One cannot consistently claim that (1) volitionism and RRV both rely on the same kind of control, (2) this kind of control is robust enough to warrant the reactive attitudes, but (3) the control to which RRV appeals is not robust enough to warrant the reactive attitudes. Volitionism and RRV must appeal to substantially different types of control. But second, and more importantly, this objection requires a good deal of support. Yet volitionists rarely offer this support in favor of the objection. It is unfair to accept that RRV appeals to a different type of control and then reject RRV precisely because it appeals to a different type of control without showing that this different type of control is insufficient for responsibility. After all, Angela Smith thinks this rational control grounds our responsibility for attitudes and responses in a meaningful and robust way (Smith 2008: 386). So what case is there to be made that rational control is too weak to ground responsibility?

Holly Smith helpfully unpacks the objection on behalf of the volitionist.75 We might think of two different types of psychological processes:

\[ P1: \text{desire} \rightarrow \text{act process} \]

75 Holly Smith herself rejects this volitionist line of argument.
In the first process, an agent’s mental state (e.g., a desire-belief pair, or an intention) causes an act (e.g., raising a hand). In the second process, an agent’s mental state causes another mental state (e.g., a failure to notice some feature, or a pang of dismay) or a bodily state (e.g., blushing) (H. Smith 2011: 122). The upshot of P1 is a prototypical example of the sort of item for which agents are responsible, according to the volitionist. The upshots of P2, on the other hand, will often involve nonvoluntary items for which the volitionist denies agents are (directly) responsible. The volitionist’s claim is that only P1 is sufficient to ground responsibility in the sense that warrants the reactive attitudes. In order to defend this claim, the volitionist must show what relevant difference there is between P1 and P2 such that the upshot of P1 justifies the reactive attitudes while the upshot of P2 does not.

As I pointed out above, volitionists rarely offer such a defense themselves. Smith offers three on their behalf, but argues that none of them survive scrutiny. First, a volitionist might claim that in P1 the agent can resist performing the desired act, but in P2, the agent can’t resist engaging in the response process caused by some attitude (123). Smith argues that this defense fails for two reasons. First, it doesn’t seem to be the case that every act that is the result of P1 is the sort that can be resisted. Many of us can relate to having a desire not to eat a rich dessert for health reasons. Nevertheless, we fail to resist the desire and eat the dessert. Of course, if such a desire is irresistible one might claim that the agent isn’t responsible for the resulting act precisely because the agent had no capacity to resist the desire.

But if a volitionist claims that agents are responsible for the results of processes that can be interrupted, it looks like some nonvoluntary, unintentional responses might also get in the door. Agents can sometimes resist the responses

---

Smith uses desire as the initiating component of P1. One might think it would be better to use intention instead, especially given that I understand the volitionist thesis as VT2. For ease of discussion, I use desire, rather than intention, as Smith does. Plausibly, some sort of desire would be relevant in forming an intention, so using desire here shouldn’t be a problem.

P2 begins with an attitude rather than an evaluative judgment, but I follow Holly Smith’s formulation here because I understand it to be the broader of the two. Holly Smith often refers to attitudes issuing in some response, while Angela Smith often refers to judgments issuing in some attitude. Because Angela Smith is particularly interested in capturing our responsibility for attitudes, she focuses on the judgments issuing in attitudes. But given that the attitude can itself be an evaluative judgment and the response process can itself be an attitude, the broader process in P2 adequately captures Angela Smith’s view. I return to and expand upon this point below.

This may be because some other desire was stronger than the original desire, or perhaps because some second-order desire was stronger than the first-order desire. In either case, it isn’t clear that the act that results from the desire can be resisted (see Smith 2011: 123).
caused by their attitudes—even if these responses are nonvoluntary. Smith offers several examples of nonvoluntary responses stifled or interrupted via the attitude → response process: “restraining a scream under torture, maintaining a poker face during stressful situations, or suppressing a wince when receiving a flu shot” (123). Both P1 and P2 seem to have instances in which the process can be interrupted as well as instances in which the process cannot be interrupted. So the difference between P1 and P2 can’t lie in the fact that one process, but not the other, cannot be interrupted.

One might object that each of these examples results in some behavioral response, but RRV is meant to capture attitudinal responses as well. If the upshot of P2 is coming to have some attitude as the response process, it may well seem that such an upshot cannot be resisted. Consequently, the fact that the attitudinal upshot of P2 cannot be resisted while the upshot of P1 can sometimes be resisted may well be a morally relevant difference.

This objection might be clearer when fleshed out. Pamela Hieronymi has examined the claim that “blame has a characteristic force, a force which is not fairly imposed upon the wrongdoer unless certain conditions are met—unless, e.g., the wrongdoer could have done otherwise, or is able to control her behavior by the light of moral reasons, or played a certain role in becoming the kind of person she is” (2004: 115). As I read the volitionist charge here, the reason why this ability to resist the upshot of P1 but not P2 is apparently morally relevant is that it makes a difference to whether one can be justified in holding another responsible, say by blaming the other person. The thought is that one cannot be so justified, because blame has a special force, and it would be unfair to subject one to that force if one could not have resisted the upshot for which one is supposedly blameworthy.

If this is the volitionist objection (and I believe it is), then the nonvolitionist might accept that the ability to resist the upshot of a process is a difference between P1 and P2, but reject that it is a morally relevant difference. Hieronymi argues convincingly that despite its characteristic force, blame is not unfair even if an agent could not have done otherwise. If Hieronymi is correct that blame is not unfair in these cases, then it does not seem that the fact that the upshots of P1 can be resisted while the attitudinal upshots of P2 cannot is a morally relevant difference.

The reason that blame is not unfair in these cases, according to Hieronymi, lies partly in the nature of blame. Hieronymi holds, as I do, that blaming someone involves, at least partly,

---

79 I owe this objection to Randolph Clarke.
adopting certain reactive attitudes toward them or making certain judgments about them. These reactive attitudes are reactions to judgments about some agent’s quality of will, or how they regard other members of the moral community. They consist in some negative affect as well as some judgment, such as the judgment that one has not been shown the due regard one warrants as a member of the moral community. According to Hieronymi, the force of blame lies in these judgments that largely constitute it, because it matters to us whether certain judgments are true, and it matters to us whether others think such judgments are true (122). Thus, “a judgment gains force from the importance of its content and the importance of the opinions of others on that topic” (122).

It seems that a judgment of a lack of due regard is fairly made when one is justified in inferring that an agent has a lack of due regard for another. One is justified in making this inference when one has adequate reasons to think that an agent has a lack of due regard. But one who claims that blaming is unfair does not maintain that there was not a lack of due regard or that one is not justified in so inferring this lack of due regard. The charge is that despite the justification of this judgment, the force of the judgment is unfair. However, the interpersonal significance of the judgment has no bearing on the reasons that justify it as being accurate. That is, one can be justified in judging that a lack of due regard has been shown regardless of the force of that judgment, because the force does not bear on the reasons for thinking that due regard has been shown. Hieronymi concludes, “Thus the force of the judgment cannot render the judgment unjustified, and so cannot render so judging unfair” (129).

Yet given the force of these judgments, one might think that even if some judgment is fairly made, one should only make this judgment when it is deserved. Whether a judgment is fairly made does not determine whether the judgment, and consequently the force of that judgment, is deserved, as Hieronymi acknowledges. An inaccurate but fairly made judgment of ill will carries a force that is undeserved in virtue of its inaccuracy. An accurate but unfairly made judgment carries a force that might nevertheless be deserved, as when someone makes a hasty judgment that happens to be true. Given this possibility, she writes, “[W]e should allow a gap between fairness and desert, to track the gap between justification and truth. It seems, at least at first, that the judgment is fair if justified but deserved if true” (130). If this is right, then the force of the judgment is deserved where its content is accurate. Yet this charge of unfairness
should grant that the judgment can be both justified and true, so the force of the judgment may be neither unfair nor undeserved.

These judgments that partially constitute the reactive attitudes are accurate when one has, in fact, been shown ill will or has not been shown due regard (in the case of blameworthiness). No further conditions, such as the ability to do otherwise, need be met for the judgment to be accurate or justified. And if the judgment is accurate and justified, then the force of the judgment is neither unfair nor undeserved. So the fact that one’s values and judgments result in certain attitudes that cannot be resisted, unlike some act process that results from an intention, does not imply that it is unfair to blame the agent or that the agent does not deserve blame. This difference between P1 and P2 is not morally relevant.

To be sure, this is a rather staunch Strawsonian compatibilist reply. But note that if it is the ability to resist some upshot or to do otherwise that the volitionist insists makes a morally relevant difference, it is important to recognize that even among volitionists, there is a dispute regarding whether the fact that one could have done otherwise than she did is a difference relevant for moral responsibility. Semicompatibilists, like John Martin Fischer, insist that it is the actual sequence of events that is relevant for moral responsibility, as the actual sequence reveals something significant and morally assessable about the agent even if the agent had no alternative possibilities. Yet Fischer seems to subscribe to the thesis that one can be directly responsible only for one’s actions and simple omissions. If volitionists appeal to the resistibility of some upshot as a morally relevant difference between P1 and P2 such that one can be responsible for upshots of P1 but not P2, then volitionists threaten to isolate a significant portion of volitionists who might reject this reasoning.

Another possible justification is that because the initiating mental states in P1 are voluntary, they transmit some special voluntary status to the actions they cause. Attitudes, however, are different; they are not voluntary and so don’t transmit this special status to their responses (Smith 2011: 123). But, Smith responds, desires and intentions don’t themselves seem to be voluntary, so this justification rests on a mistake. Whether one has some desire or intention is not always up to one, nor is it something one chooses to have. Many intentions are

---

80 In particular, there is a debate over whether Frankfurt-style cases show that being able to do other than one does is required for one to be morally responsible. Several volitionists accept that Frankfurt-style cases are successful. See, for instance, Fischer and Ravizza 1998.
81 Hieronymi 2008 argues this point persuasively.
non-actively acquired, as we saw in chapter 2, and do not seem to count as voluntary (to the extent that we can understand what volitionists mean by “voluntary”). Yet the actions that result from these non-actively acquired intentions are still intentional, and I suspect that volitionists will still claim that such intentional actions are voluntary. If so, there are sometimes upshots that are voluntary even though their initiating intentions or desires are not voluntary, so this voluntary status is not always transmitted in P1. Consequently, a reliable difference between P1 and P2 cannot be that the initiating mental states in P1 transmit some voluntary status to their upshots whereas the attitudes in P2 do not transmit such a voluntary status to their upshots.

Lastly, the volitionist might insist that blameworthiness requires rationality, but only intentions have this feature of rationality. After all, intentions are sensitive to new information. One might intend to put some substance she believes to be sugar in her coffee, but upon learning that the substance is actually arsenic, the rational person would no longer intend to put the substance in her coffee (123). Again, Smith argues, these nonvoluntary attitudes involved in P2 are also often sensitive to new information in this way: “one’s admiration for a colleague may dissipate on learning that she has been defrauding the company, and one’s loyalty to a friend may increase on learning that he has been quietly working to support one’s career advancement” (123). This connection with rationality is one of the benchmarks of RRV, so this justification also fails to draw a distinction between P1 and P2 in the way the volitionist desires.

It doesn’t seem that the volitionist who insists that there is an important and relevant difference between P1 and P2 can explain either what this difference is or why it would make a difference to grounding responsibility. Upon further inspection, the two types of processes seem more closely related than at first blush. I claimed above that evaluative attitudes can give rise to voluntary responses, but similarly, desires and intentions can sometimes give rise to nonvoluntary responses. As Smith explains, one’s intention to eat the last donut may cause one to reach for the donut. But it might also cause nonvoluntary responses, such as thinking about the donut, salivating, and anger if someone else takes the donut first (124). The fact that the initiating elements in both P1 and P2 can result in both sorts of upshots makes it difficult to see why there is a relevant difference between the two types of processes.

The salient difference between P1 and P2 seems to be one of intentionality: in P1 the intention involves the aim to do something, which gives rise to that effect (124). An intention normally causes the relevant action that is part of the action plan of that intention. In contrast,
feeling embarrassed may cause one to blush, but the blushing isn’t an aim or a part of being embarrassed. So paradigmatically, the initiating element of P1 involves some aim, and the upshot of P1 involves satisfying that aim. But also paradigmatically, the initiating element of P2 does not involve some aim, and the upshot of P2 does not involve satisfying some aim (125). But is this a morally relevant difference such that P1, but not P2, can result in blameworthy upshots?

Smith argues that it is not. In fact, this aiming is an important feature of why an act can be considered voluntary. Thus:

Unless you take ‘blameworthy’ to entail ‘voluntary’ it doesn’t seem to follow directly from the fact that these responses are non-voluntary (something admitted up front by Attributionists) that the agent is not responsible for them and not blameworthy if the underlying attitude is morally objectionable. Of course the Attributionist denies any such entailment. (2011: 125-126)

Ruling out the upshots from P2 as not blameworthy in virtue of the upshots not fulfilling some intention or aim begs the question against the nonvolitionist. One cannot simply rule out such nonvoluntary responses from the realm of responsibility because they are nonvoluntary. This seems the volitionist’s only recourse, however. Consequently, there is no justification for claiming that upshots from P2 cannot be things for which we are directly responsible in the sense that warrants the reactive attitudes.

5.3.3 Too Wide

According to RRV, we are directly responsible for our attitudes and other mental states because they are normatively connected to our judgments. Many nonintentional mental states, including physical sensations and pains, hunger and thirst, and the like are not normatively connected to our judgments. As Angela Smith writes, “we do not expect these states to be rationally sensitive to our evaluative judgments or our wider cognitive and evaluative commitments” (2005: 257). Consequently we are not directly responsible for these states on RRV. But Smith also acknowledges that some of our physical states are directly connected to our evaluative judgments in the requisite way. To use Smith’s example, I may feel nauseous before speaking publicly. This nausea “seems to be a direct result of my evaluative judgment that
such public performances are both important and also fraught with opportunities for failure” (257-258). But we do not normally think that agents are responsible for nausea before speaking publicly. Insofar as RRV holds that agents are directly responsible for physical states for which we do not think agents are responsible, the view captures too much. Either Smith must accept this wider view as a revisionist cost of her theory or she must explain why agents aren’t responsible for these physical states on her view.

Smith grants that we are responsible for our physical states like nausea before public speaking, but argues that this responsibility is different from our responsibility for, say, our attitudes. The connection between our evaluative judgments and these physical states is of a different sort than the connection between our evaluative judgments and attitudes like shame or fear. In the case of these physical states, there is merely a causal connection between the judgments and the physical states. The judgment causes some physical state, but the physical states aren’t sensitive to judgments. She writes, “For that reason, the kind of responsibility we have for them is essentially derivative: it flows from the responsibility we have for the evaluative judgments which constitute their causal triggers” (258). No one could criticize us for failing to feel nausea even though we make the evaluative judgment that public performances are important and may result in failure.

In contrast, however, Smith claims that our attitudes are not merely caused by these evaluative judgments: “They are, rather, active states, in the sense that they essentially involve our judgmental activity. So our responsibility in these cases is direct and normative: we have a responsibility for the internal rational justification of the attitude itself” (258). For example, feeling shame is not only caused by a certain judgment (say, that one has done wrong or is disgraceful in some way). Shame fundamentally involves that judgment. One could be criticized for failing to feel shame when one makes the evaluative judgment that treating someone a certain way is wrong and one has treated a person that way.

---

82 One related worry is whether we are responsible for the random thoughts and attitudes that occur to us throughout the day. Various marketing slogans or jingles are designed to stick in our minds, but Smith claims that these sorts of random thoughts and attitudes are not connected to our evaluative judgments in the right way. We could not ask for justification for some mental image, song lyric, or the like, according to Smith (2005: 260). But she acknowledges that there are fuzzy boundaries here. It is not so clear that when someone is repeatedly struck by violent or pornographic images, these images reveal nothing about his values and commitments. If there is a pattern, this pattern does seem to reflect an underlying judgment that these things are important. And if there is some pleasure associated with these images, this also is an indication of some underlying evaluative judgment. But, according to Smith, “The more random and isolated these occurrences are, however, the less plausible it may seem to say that they reflect a person’s judgments and evaluative commitments” (2005: 261).
Smith’s explanation, then, is that our responsibility for these physical states is derivative because the judgment merely causes the physical state. But our responsibility for our attitudes is direct because the judgment not only causes the attitude, but our attitude is constituted by our judgment.

This response sits uncomfortably with another commitment that Smith has, which Holly Smith flags (2011: 122, n. 20). Angela Smith holds that agents are directly responsible for neglecting, forgetting, or failing to notice certain things. Yet in these cases, the nonvoluntary response is not constituted by a judgment, but merely caused by it. For instance, in Birthday, Angela’s failure to call Dave on his birthday is not constituted by her judgment that fulfilling promises is unimportant or that Dave is unimportant. If Angela makes such a judgment, this judgment can merely cause her failure at best.83

The treatment of responsibility for physical states and responsibility for omissions, forgetting, and cases of neglect must be the same. While Holly Smith does not press the point, the problem for RRV seems to be the following: Either agents are only derivatively responsible for omitting, neglecting, forgetting, or failing to notice something in virtue of a mere causal connection between the judgment and the response, or else agents are directly responsible for these items. If agents are only derivatively responsible for omitting, neglecting, forgetting, or failing to notice something, then Angela Smith must revise a portion of her view. RRV currently holds that agents are directly responsible for such items. If agents are directly responsible in these cases, however, then there seems no principled way in which Angela Smith can distinguish between responsibility in these cases and cases of physical states such as nausea before public speaking. In this case, Smith must hold that the judgment’s causing the response is sufficient for direct responsibility.

I do not see how Smith can plausibly adopt the second line of response. It seems that agents are only derivatively responsible for physical states, omissions, negligence, forgetting, and failing to notice on any plausible rendering of RRV. One can see the derivative nature of responsibility here by asking whether an agent would be responsible for some omission or physical state if one were not responsible for the attitude that caused it. It seems that the only

---

83 This requires more careful treatment. If Angela’s omission is nothing, as many omissions are, then it seems that her judgment cannot cause the omission (see chapter 3). But her judgment may cause her subsequent behaviors, and there may be a way to paraphrase this relation between her judgment and her omission. It seems far less plausible, however, that Angela’s judgment could constitute her omission when that omission is nothing at all.
reason an agent is responsible for some physical state or omission is because the agent is responsible for the attitude that caused the omission or physical state. Thus, one’s responsibility for the omission or physical state is derivative and traces back to one’s responsibility for the attitude that caused it.

Contrast this derivative responsibility with an agent’s direct responsibility for attitudes. As Smith argues, many of our attitudes are constituted by some judgment or other, so there is no tracing involved in such cases. Responsibility for these attitudes is direct.\textsuperscript{84} It seems, however, that responsibility for anything else will be derivative, as it will depend on our responsibility for our attitudes. If an agent is not responsible for her attitudes, then the agent cannot be responsible for the actions, omissions, or physical states that are caused by these attitudes. This is the nature of derivative responsibility.

I think, then, that the first line of response is the most plausible. We are only derivatively responsible for physical states, but we are also only derivatively responsible for omissions, neglecting, forgetting, and failing to notice. One might raise two concerns with this response. First, perhaps it is a problem that agents are only derivatively responsible for omissions, etc. because this derivative responsibility is inferior to direct responsibility. Second, even if agents are only derivatively responsible for physical states, they are nevertheless responsible for those states. That consequence of RRV may seem to some too implausible to accept. I take each of these concerns in turn.

\textsuperscript{84} One might think that responsibility for these attitudes that are only partly constituted by judgments cannot be direct either. Suppose some attitude is partly constituted by some judgment. Using the test for derivative responsibility, we can ask whether an agent would be responsible for this attitude if she were not responsible for the judgment. It seems plausible that she wouldn’t be responsible for the attitude. Nevertheless, as Smith seems to understand direct responsibility, one can be directly responsible for attitudes that are just partly constituted by some judgment, and so can be directly responsible for some item even if one is not directly responsible for all of that item’s constituents.

One might reject this understanding of direct responsibility. Fortunately for Smith, this understanding of direct responsibility is not required for her to be able to draw the distinction between responsibility for attitudes, noticing, neglecting, and forgetting on the one hand, and responsibility for physical states on the other. Rather than appealing to constitution, Smith might appeal to rational assessability. Consider again one’s nausea before public speaking. The judgment that gives rise to nausea before public speaking does not recognize reasons for nausea. Instead, one’s judgment that public speaking is dangerous and carries the potential for failure gives rise to the nausea. This is different from a case of failing to notice something, for example. The judgment that something is unimportant is sensitive to reasons for not noticing it. There is a kind of normative connection between the not noticing and the judgment that does not hold in the case of nausea. There aren’t reasons for or against nausea or other such physical states, so these states aren’t rationally assessable. This is so even though one of the agent’s judgments might cause the physical state (and so one could still be derivatively responsible for these physical states).
First, derivative responsibility is no less responsibility, so if an agent is derivatively responsible for omissions and the like, this is not, in itself a mark against the view. Even if derivative responsibility were somehow inferior in virtue of the distance from the agent to the item for which she is responsible, the trace from the item in question to the attitude is a very short one. Consequently, the fact that we must trace one’s responsibility back to some attitude is not at all problematic.

Given this response, however, the second concern may seem more pressing. Whether directly or derivatively, agents are still responsible for their physical states. If derivative responsibility is just as robust as direct responsibility, it seems that Smith’s response doesn’t go far enough in assuaging concerns that it is quite odd to say that agents are responsible for physical states. While I am sympathetic with this concern, it’s important to note several things that may make this claim seem less incredible than at first blush.

Perhaps most crucially, RRV does not entail that agents are responsible for all of their physical states, but only those that are normatively connected to judgments in the right way. Thus, only physical states that are sensitive to our evaluative judgments or commitments would count. Perhaps blushing after a bumbling romantic encounter reflects a judgment that one’s behavior is shameful or embarrassing. Nausea before a long road trip in a U-Haul may reflect a judgment that such a trip is dangerous. Any state that is responsive to evaluative judgments—one for which one might be asked for a defense or justification—is one for which an agent may be responsible. Of course, states like being hungry, thirsty, or in pain, as well as reflexes, do not seem to fit this criterion. They are not rationally sensitive to our evaluative judgments, and it doesn’t make sense to ask for a justification for these states.85

Additionally, I think this claim sounds implausible in large part because of the close connection between responsibility and blameworthiness. In order to be blameworthy for some item, however, there must be some moral norm that the agent violates. It doesn’t seem that there is any such norm in these cases of physical states. For instance, it is false that one morally ought to feel nauseous before speaking publicly if one is nervous and judges the opportunity to be harrowing. It is certainly understandable, and perhaps it is rational if the judgment is well-founded. The agent might be criticized as irrational for not having butterflies in her stomach before speaking if she judges the event as full of opportunities for failure. But this is quite

different from saying that one ought to believe that harm is bad or that people deserve respect, and quite different from the criticism one might be rightly subject to for failing to hold these beliefs. So while one may be responsible for the physical state in virtue of its rational connection to one’s judgments, it is an independent question what sort of treatment one deserves on the basis of this responsibility. If it is not the case that one ought to experience some physical state in response to some judgment, then one might be morally responsible but not morally criticizable for that state.

Even if this is mistaken, however, and agents are blameworthy for their physical states, it seems unlikely that one deserves a high degree of blame for these states, as they are of minor moral significance. Additionally, even if one is blameworthy for some physical state, it is a further question of whether anyone in particular ought to blame the agent for some physical state. Blameworthiness is but one defeasible consideration in favor of blaming an agent, but there may be many overriding considerations that defeat this consideration. The fact that many of us feel nausea before public speaking might undermine our standing to blame another person for feeling nausea in this case. Add to this that there is little good that comes from blaming an agent for her nausea and one’s blameworthiness for the nausea may not translate into any actual blame for the nausea. This lesson can be expanded to other physical states, so that even if agents are blameworthy for their physical states, it is plausible that anyone rarely, if ever, ought to actually blame agents for such states. Thus, even if agents are responsible for their physical states on RRV, this is not a result of drastic import or one that should cause us to abandon the view altogether.

5.3.4 Too Narrow

Although RRV has been criticized as being too wide, some writers have also criticized the view as being too narrow. They claim that the view fails to capture an agent’s responsibility for some nonvoluntary fault in several cases, such as certain forgetting cases, failure to notice cases, and negligence cases. RRV holds that an agent is responsible for some nonvoluntary item only insofar as it is connected to an evaluative judgment in the right way. And in some cases, an agent’s failure to notice a certain feature may reflect a judgment that the relevant feature is not significant. The problem, however, according to these objectors, is that some of these

---

86 I’ll return to this idea of degrees of blameworthiness and punishment for nonvoluntary items in chapter 7.
nonvoluntary failures are not judgment-based. The view cannot explain our responsibility in these cases, they claim, but any plausible view should explain this responsibility.

Consider the following two cases offered by George Sher (2009: 24):

**Hot Dog:** Alessandra, a soccer mom, has gone to pick up her children at their elementary school. As usual, Alessandra is accompanied by the family’s border collie, Bathsheba, who rides in the back of the van. Although it is very hot, the pick-up has never taken long, so Alessandra leaves Sheba in the van while she goes to gather her children. This time, however, Alessandra is greeted by a tangled tale of misbehavior, ill-considered punishment, and administrative bungling which requires several hours of indignant sorting out. During that time, Sheba languishes, forgotten, in the locked car. When Alessandra and her children finally make it to the parking lot, they find Sheba unconscious from heat prostration.

**On the Rocks:** Julian, a ferry pilot, is nearing the end of a forty-minute trip that he has made hundreds of times before. The only challenge in this segment of the trip is to avoid some submerged rocks that jut out irregularly from the mainland. However, just because the trip is so routine, Julian’s thoughts have wandered to the previous evening’s pleasant romantic encounter. Too late, he realizes that he no longer has time to maneuver the ferry.

Sher judges that both Alessandra and Julian are responsible in these cases. Yet we cannot trace back responsibility to some prior failing. Alessandra always brings Bathsheba with her to pick up her children, and she had no reason to suspect that this pick-up would be different from others. She could have set an alarm on her phone in case she wasn’t out of the school in some short amount of time. But again, she had no reason to suspect this pick-up would be any different, and it doesn’t seem like she is obligated to set some alarm in every other case. Similarly, Julian’s trip is routine. He has no reason to suspect that his thoughts will wander, we might stipulate. He is normally focused on his work. There’s no reason to think that he should have written himself some reminder, for instance, that he should focus on piloting the ferry. Julian doesn’t choose to let his mind wander, nor does he intentionally let his mind wander. He simply finds himself...
struck with some distracting thoughts on this particular occasion. So it seems implausible that Alessandra and Julian are derivatively responsible for their respective failures. If they are responsible at all, they must be directly responsible.

According to RRV, if Alessandra is directly responsible for wrongdoing, it is in virtue of some judgment that is expressed by her forgetting Sheba in the car. Sher offers possible candidates: Alessandra judges that Sheba’s safety isn’t very important, or Alessandra lacks some degree of good will toward Sheba. Similarly, Julian would have to be responsible in virtue of some judgment that the safety of his passengers isn’t terribly important, or that focusing on what he is doing is not worth his time or effort. Sher responds:

[A]lthough it is certainly possible to suppose that Alessandra and [Julian] have made such judgments, their responsibility does not appear to depend on this. Alessandra would surely remain responsible if she cared a lot about Sheba but was simply distracted by the volume and intensity of the dispute.... (2008: 226)

Similarly, Sher thinks that Julian would remain responsible even if he made no such judgment.

Perhaps more to Sher’s point, however, it is more plausible that Alessandra made no such underlying judgment in the first place. According to Angela Smith:

[I]f one judges some thing or person to be important or significant in some way, this should (rationally) have an influence on one’s tendency to notice factors which pertain to the existence, welfare, or flourishing of that thing or person. If this is so, then the fact that a person fails to take note of such factors in certain circumstances is at least some indication that she does not accept this evaluative judgment. (2005: 243)

So the fact that Alessandra fails to notice certain features of her situation, such as Sheba being in a hot car, is some indication that she doesn’t accept the evaluative judgment that Sheba’s welfare matters much. But, Sher contends, “the urgency of the dispute, and its high emotional volume, seem by themselves to be quite sufficient to explain her failure to notice her act’s wrong-making
feature” (2009: 131). We need not appeal to some implicit or explicit judgment that Alessandra makes to explain her cognitive failure in this case. And, Sher claims, there are likely a great many cases in which this is true. We need not infer that Julian has made a judgment that the welfare of those on the ferry doesn’t matter much; the mundane trip and the pleasant encounter from the previous night might be sufficient to explain why he fails to avoid the rocks.

If Sher is correct, then RRV is insufficient to explain our responsibility for several nonvoluntary faults. RRV cannot explain Alessandra or Julian’s responsibility without an implausible or unnecessary claim that each held some objectionable evaluative judgment that explains their failure. But it seems that neither Alessandra nor Julian made explicit or implicit judgments that constituted or caused their attitudes or their nonvoluntary responses.

Similar objections against RRV and its ilk have been raised by Matt King (2009: 583-586). According to King, in many cases of negligence the agent does not express any objectionable evaluative judgment, because the agent does not make any such judgment. When Nate backs out of his driveway in the morning without checking behind his car, distracted by the radio and thoughts of work, he does not express any objectionable evaluative attitude such as, “The welfare of anyone behind my car is unimportant to me.” He is simply distracted, and this distraction is sufficient to explain his negligence. Yet we still want to blame him for hitting a child when he backs out of the driveway.  

Holly Smith further highlights the difficulty of failing to capture all our intuitions in these sorts of cases in which there is no objectionable evaluative attitude:

Nor does it appear that Attributionism [e.g., RRV] has any resources for explaining why we may feel that the agents in these original cases are blameworthy. Even if some independent explanation can be found, the prospects look bleak for achieving anything like a unified account of what makes agents blameworthy in non-tracing failure-to-notice cases. (2011: 120)

---

87 Levy 2011 makes a similar point. He argues that just because an agent fails to notice something or forgets some promise or obligation, we cannot infer that the agent does not care sufficiently about the item in question. Many distracting factors might be sufficient to explain the lapse or failure. I’ll have more to say about this in 7.3.2.

88 I discussed this case in chapter 4.
This is especially important in light of the criticism I advance against volitionists. Just as volitionists can’t appeal to an ad hoc justification for including choice as voluntary or for adding intentional omissions in VT2, nonvolitionists can’t appeal to some independent account to explain these difficult cases without being similarly ad hoc.

Angela Smith has responded to some of these criticisms—most notably Sher’s. Before examining her response, however, it will be instructive to examine Sher’s own account of responsibility to see why these sorts of objections aren’t as troubling for the proponent of RRV as they initially seem.

5.4 Wider Views and the Grounding Problem

If RRV cannot explain why Alessandra and Julian, among others, are responsible for their unwitting behavior, what can explain this responsibility? Sher offers his own account of why agents need not be aware of the wrongfulness of their behavior in order to be responsible for that behavior. Whereas Smith requires a rational connection between an evaluative judgment and some attitude, Sher requires something much weaker.

For Sher, one can be morally responsible for unwitting behavior because one’s failure to realize this wrongful behavior is caused by a combination of the agent’s constitutive attitudes, dispositions, and character traits:

[W]hen an agent should, but does not, recognize that he is acting wrongly or foolishly, what connects him to the act’s wrongness or foolishness in a way that allows us to hold him responsible is not just his failure to live up to whatever standard requires that those in his position recognize such acts as wrong or foolish, but is rather the whole collection of attitudes, dispositions, and traits whose interaction causes him not to recognize this. (Sher 2009: 87)

These agential aspects that make a particular agent the person that he is cause the failure, and for Sher this causal connection can explain why an agent is responsible for that failure.

This causal connection is but one part of a much larger epistemic condition that Sher offers (143):
**FEC:** When someone performs an act in a way that satisfies the voluntariness condition, and when he also satisfies any other conditions for responsibility that are independent of the epistemic condition, he is responsible for his act’s morally or prudentially relevant feature if, but only if, he either

1. is consciously aware that the act has that feature (i.e., is wrong or foolish or right or prudent) when he performs it; or else
2. is unaware that the act is wrong or foolish despite having evidence for its wrongness or foolishness his failure to recognize which
   a) falls below some applicable standard, and
   b) is caused by the interaction of some combination of his constitutive attitudes, dispositions, and traits; or else
3. is unaware that the act is right or prudent despite having made enough cognitive contact with the evidence for its rightness or prudence to enable him to perform the act on that basis.

Because the cases with which I am concerned are primarily those of unwitting wrongdoing, I will focus on the second condition of FEC.

Without 2a, agents would nearly always be responsible for unwitting wrongdoing, since one’s failure of awareness is almost always caused by the interaction of features that constitute the agent. On Sher’s account, this applicable standard is relativized to some particular agent’s cognitive capacities and the moral requirements that apply to him (113). It is not, for instance, determined by what most agents would be aware of in the relevant situation. So Alessandra is responsible for forgetting Sheba not because most people would remember Sheba in the car, or a reasonable agent would remember Sheba, but because forgetting Sheba fell below the threshold.

---

89 I use “unwitting” rather than “nonvoluntary” here because FEC is focused on an epistemic condition and supposes that an agent fulfills some voluntariness condition on responsibility. There are complications regarding whether an agent can behave voluntarily without being aware of the relevant features of her behavior. Sher understands this problem, but also proposes that the voluntariness condition on responsibility be understood in a way that “requires neither awareness nor control” (149). Most would think that the voluntariness condition must require at least one of these, which suggests that Sher is offering an account more in line with nonvolitionist accounts (though, as I argue in chapter 2, our notions of voluntariness and voluntary control are not clear). To complicate matters, Sher seems to think that agents are only responsible for actions or omissions, but he doesn’t seem to hold that agents are directly responsible only for intentional actions or omissions, as Julian, for instance, does not intentionally fail to avoid the rocks in On the Rocks. Whether Sher’s view should properly be labeled a volitionist view or a nonvolitionist view is a difficult question, but because nothing substantial hangs on the label, I set aside the problem here.
of Alessandra’s cognitive abilities (and Alessandra was obligated to care for Sheba, which required remembering her in the car). In this way, Sher limits exactly when an agent might be blameworthy for some unwitting behavior while connecting the agent with the wrongdoing in a more intimate way.

While 2a limits responsibility, 2b explains how an agent is responsible for unwitting behavior. Sher argues that a responsible agent should be identified with his consciousness or reasons-responsiveness, but also with “whatever psychological and physical structures sustain his normal patterns of intellectual functioning” (122). For Sher, a person’s background beliefs, moral commitments, views about what is valuable, degree of optimism or pessimism, attitude toward risks, and emotional makeup are constitutive of the person as an agent (122). For example, we can attribute Alessandra’s failure to remember Sheba to her “concern for her children, and her tendency to focus intensely on whatever issue is at hand” (123). Notice that these constitutive features need not be bad or flawed features; they may involve bad features, good features, or a mix (91). Sher writes (92):

Even if none of these traits or attitudes are bad, it remains true that if they were different enough—if, for example, Alessandra were less solicitous of her children, or was made less anxious by conflict—then she would not have forgotten about the dog. Thus, in Hot Dog, no less than in the bad-character cases, we can make sense of the agent’s responsibility by taking Alessandra herself—the whole person—to be the source of her own failure to draw the connection between what she knows and the wrongness of her act.

These constitutive features of the agent explain why the agent failed to meet some obligation that she had, and for Sher, this is sufficient (along with his other conditions) to explain the agent’s responsibility for her unwitting bad behavior. It is sufficient to explain why Alessandra is responsible for forgetting Sheba, and it requires no implausible appeal to Alessandra making some objectionable evaluative judgment about Sheba’s welfare.90

---

90 Sher’s view is unique, but Randolph Clarke offers a related view. He proposes the following sufficient condition for direct blameworthiness for a wrongful, unintentional, unwitting omission:

Provided that the agent has the capacities that make her a morally responsible agent, she is blameworthy for such an omission if she is free in failing to doing [sic] the thing in question and if her lack of awareness of her
Sher’s responsibility without awareness view (RWA) is complex, and it takes into account important information about what constitutes an agent, what norms are at issue in holding an agent responsible, and how to properly connect an agent with her behavior so as to capture our everyday judgments in responsibility ascriptions. Is RWA more viable than RRV?

I argue that it is not. RWA does not seem to offer sufficient grounding for an agent’s responsibility. The general problem is that while Sher is keen to show that there is an important connection between the agent and her behavior, “merely citing a causal connection between some failure of awareness and the workings of the vast ‘psychophysical structure’ that generally sustains our intellectual activities cannot establish the right kind of connection between an agent and her wrongdoing to justify us in regarding her as responsible for it” (Smith 2010: 523).

According to RRV, an agent’s evaluative judgment is often reflected in her behavior. This judgment, for which we can demand justification, is what grounds an agent’s responsibility. But on RWA, no such judgment is necessary; the agent is responsible if her constitutive features caused her to act in a way that fell below an applicable standard. On RWA, it is difficult to see why Alessandra is responsible for forgetting Sheba in the car. Assuming Alessandra’s forgetting does not reflect any objectionable evaluative judgment, we are left with Sher’s appeal to Alessandra’s constitutive features causing the lapse. But Matthew Talbert puts the point nicely when he writes, “[M]erely noting that Alessandra’s lapse was caused by internal psychological and physical structures does not add much to simply stipulating that Alessandra forgot about Sheba” (2011: 150). And restating the problem—that Alessandra forgot about Sheba—doesn’t give us any further grounding.

Perhaps Sher might insist that we should focus on capacities instead as those who favor capacity views and VT3 do. Certain standards apply to Alessandra because of her capacities. She could have been aware of the wrongness of her behavior, given her psychological constitution. And she was obligated to care for Sheba, so she should have been aware of the wrongness of her behavior. But this appeal to Alessandra’s capacities, even when considering that they apply only to her, is unsatisfying. I dismissed capacity views that do not appeal to a quality of will condition obligation to do it—and of the fact that she isn’t doing it—falls below a cognitive standard that applies to her, given her cognitive and volitional abilities and the situation she is in. (2014: 167)

Clarke’s view is more restrictive than Sher’s in the sense that he is focused on a certain type of omission. Nevertheless, like Sher, Clarke does not think that the cognitive standard is a moral obligation or an ideal (167). Similarly, Clarke’s cognitive standard is relative to the agent, as Sher’s is. Because Sher’s account is broader in scope, I focus on that view over Clarke’s view. But some of what I say below in response to Sher’s view is applicable to Clarke’s view as well.
in chapter 3 in part because these views failed to offer sufficient grounding for an agent’s responsibility. The mere fact that an agent had the ability to intentionally act or omit is not enough to explain why the agent is responsible for what she actually does. If that reasoning is compelling, it should be similarly compelling here. Pointing out that Alessandra had the capacity to realize the wrongness of her conduct provides no help in explaining why she is responsible for actually conducting herself in that way.

If this reasoning isn’t enough to dismiss RWA, we might also consider whether the constitutive features to which Sher appeals are the sort to which we would appeal when we hold Alessandra responsible. Neil Levy 2011 thinks that RWA faces difficult problems here—especially when Sher allows that an agent’s failure can be caused entirely by morally good features that are constitutive of her. When we stipulate that Alessandra has no objectionable evaluative judgment that is reflected in her forgetting Sheba, Sher claims that her responsibility for forgetting is explained by her “concern for her children, and her tendency to focus intensely on whatever issue is at hand” (2009: 123). But when we blame Alessandra, these aren’t the features we would cite to explain her blameworthiness. By stipulation, Alessandra is not unreasonably solicitous of her children. Instead, says Levy, we appeal to “the abnormal features of Alessandra’s situation” to explain her lapse (2011: 243-244, n. 3). The abnormal features of Alessandra’s situation might simply be outside of her intentional or rational agency and in no way reflect poorly on her. In this case, there is no good reason to hold her responsible.

In fact, in many of Sher’s cases, it is questionable whether the agents he describes are even responsible for their behavior. My tendency to blame agents like Alessandra or Julian fluctuates with the likelihood that there is any way to trace their failing to some previous behavior or that their behavior reflects some objectionable evaluative judgment. As we have seen, RRV has been criticized for failing to properly ascribe responsibility in certain sorts of cases. But the proper response for advocates of RRV to make is that without an evaluative judgment, the agent is simply not responsible. By stipulation, as Sher tells Hot Dog and On the Rocks, neither Alessandra nor Julian has any objectionable evaluative attitude to be expressed through their behavior. Also by stipulation, there is no way to trace their responsibility back to responsibility for some previous behavior. With these details filled in, it seems neither Alessandra nor Julian are responsible for their behavior.
Importantly, this is not a counterintuitive conclusion. If there is no objectionable evaluative attitude that Alessandra or Julian had, and there is no way to trace responsibility back to some previous failing, then what could possibly ground their responsibility? Sher’s answer seems to offer nothing more than a restatement that Alessandra forgot Sheba and Julian’s mind wandered, and neither agent should have done these things. But this gets us no closer to a justified story of responsibility.

Upon realizing the relevant features of Alessandra’s situation (that she expressed no objectionable judgment in forgetting Sheba and that we cannot trace her responsibility in some way), I suspect others will share my judgment that Alessandra simply isn’t responsible for forgetting Sheba. If anything, Alessandra’s forgetting Sheba in the car seems to be an unfortunate—tragic, even—situation in which no one is responsible. Alessandra might love her dog, just as many parents tragically forget their children in the car despite loving them dearly. The forgetting is more akin to a one-off glitch in her brain. If we are blaming Alessandra as Sher thinks we should, Smith writes, “Then what, exactly, are we blaming her for? …[H]ow can I hold Alessandra responsible for a mere glitch in her psychophysical system? Why isn’t this akin to blaming my computer for an uncharacteristic freeze-up” (2010: 523)?

Is the fact that RRV cannot account for an agent’s responsibility in all cases of negligence, forgetting, or failures to notice a serious cost to the view? I don’t think so. Many of these sorts of cases might involve an objectionable evaluative attitude on the part of the agent. After all, sometimes the agent is responsible for an omission or failure to notice something because the omission or failure reflects that the agent lacks certain attitudes that she ought to have. When the cases do not involve an objectionable attitude (or a lack of an attitude the agent ought to have), perhaps responsibility can be captured via some tracing technique. Nevertheless, there will undoubtedly be some cases of negligence, forgetting, failing to notice, etc. that do not involve tracing or the reflection of an objectionable evaluative judgment, and so are not captured by RRV. This is not a problem, however. When we learn that an agent did not have any objectionable evaluative judgment underlying her failure (or did not lack an attitude she ought to have had), we should also realize that we have no grounds for blaming her. A view like RWA is simply too wide. As in the case of certain capacity views, such views capture more cases at the cost of sacrificing legitimate grounding for responsibility.
CHAPTER 6

NONVOLITIONISM, CONSCIOUSNESS, AND RESPONSIBILITY

6.1 Introduction

In chapter 5, I argued that while Sher’s RWA view captures an agent’s responsibility for some unintentional or unwitting actions and omissions, it does not adequately explain why an agent is responsible for those items. RRV can offer a distinct, morally relevant justification for not only an agent’s attitudes, but also many of her unintentional and unwitting actions and omissions. Sher’s focus on awareness, however, raises important questions about consciousness and its role in responsibility.

In recent work, Neil Levy argues that to be directly responsible for an action, an agent must be conscious of the facts that give her action its moral significance. One of the central features of nonvolitionist views like RRV is that agents can be responsible for attitudes (or actions or omissions) even when those agents are unaware of the moral significance of those items. After all, one need not be aware of the moral significance of one’s forgetting a promise on a birthday or failing to notice one’s partner’s dislikes in order to reflect some morally assessable lack of care, according to RRV. So Levy’s arguments seem to stand opposed to nonvolitionist views like RRV. Despite Levy’s important and insightful arguments, I argue that consciousness

---

91 As I pointed out in chapter 1, some people may understand moral responsibility differently, which may lead to them positing different necessary conditions on direct responsibility. This is relevant because in his recent book, Levy offers a slightly different definition of responsibility: “As I use the phrase, an agent is morally responsible for an action or an omission if the fact that they have performed that action, in the circumstances and manner in which they acted, is relevant to how they may permissibly be treated when it comes to the distribution of benefits and burdens” (2014: 2). This understanding may seem too far removed from how nonvolitionists understand responsibility, as it focuses on the distribution of benefits and burdens. However, Levy later explains: “What is in question in this book is whether agents who fail to be conscious of certain facts when they act may nevertheless be blamed or praised for actions that respond to or have their content shaped by these facts. It is direct moral responsibility that is at issue…” (2014: 3). I will here assume that Levy and his opponents are only concerned with whatever kind of responsibility warrants the reactive attitudes, but I will return to the issue of responsibility and punishment in chapter 7.

92 I argued in chapter 5 that on a plausible understanding of direct and derivative responsibility, RRV should be committed to the claim that agents are directly responsible only for attitudes. While the trace to actions is a very short one, it seems agents would be derivatively responsible for their actions and omissions on such a view. If this is right, then Levy’s dispute with nonvolitionists like Smith might be different from the way he presents it in his book. Levy is concerned with direct responsibility, and he thinks that agents can be directly responsible only for behaviors,
of the facts that give some attitude or behavior its moral significance is not a necessary condition on direct responsibility. While I think Levy does give powerful reasons to think that agents cannot be directly responsible for some nonconscious attitudes given that those attitudes lie outside of our rational control, he has not shown that agents cannot be directly responsible for any nonconscious attitudes. At best, Levy’s arguments may restrict what attitudes RRV captures under direct responsibility.

6.2 The Nature and Function of Consciousness

Levy is not concerned with phenomenal consciousness, or what it feels like for an agent to experience some conscious state. Instead, he is focused on what he calls informational consciousness, or whether the agent is in a state with access to certain information (2014: 29). An agent is informationally conscious of some state when information is personally available to the entire agent (and not simply to certain parts of the mind). Levy explains that information is personally available “when the agent is able to effortlessly and easily retrieve it for use in reasoning and it is online” (33). When a state guides an agent’s behavior, it is online.

Levy holds that “in order to be morally responsible for their actions, agents must be conscious of facts that explain the valence of its moral significance, under an appropriate description” (37). This appropriate description is one that captures some aspect of the badness of the action or omission, but the agent need not be aware of every fact or description—just those facts that are sufficient to render the action or omission right or wrong (2012: 4). The claim that agents are directly responsible for some behavior only if they are conscious of the moral significance of that behavior is what Levy calls the consciousness thesis.93

Levy endorses the consciousness thesis in part because of what consciousness does in an agent. Levy is a proponent of the Global Workspace theory of consciousness (GWST), which enjoys wide support. According to GWST, the mind is composed of some number of modules, nodes, or mechanisms. Each mechanism is responsible for some direct function and is largely closed off from other mechanisms. Consciousness makes the information in one insulated

---

93 As I mention in note 92, Levy’s consciousness thesis is phrased in terms of behaviors—actions and omissions—because Levy thinks that agents can be directly responsible only for behaviors, not for attitudes.
mechanism available to other mechanisms. Essentially, consciousness allows for communication and the sharing of information from one mechanism with other mechanisms; conscious information is “broadcast…to a broad range of consuming systems (which are distinct and dissociable components of the mind)” (2014: 47). Information that is not broadcast in this way is less widely available to other mental components and not conscious. Understood in this way, the consciousness thesis holds that the facts that give some behavior its moral significance must be properly integrated into one’s global workspace such that one is aware of them.

Another reason why consciousness is important, according to Levy, is because it makes information available to the agent, and having this information available for rational deliberation is required for rational control of behavior (63). But consciousness is also required for flexibility in behavior and response. Nonconscious processes can drive complex behaviors via action scripts. An action script is “a set of motor representations, typically a chain of such representations, that can be triggered by an appropriate stimulus, and which once triggered runs ballistically to completion” (74-75). When these action scripts run without consciousness, the behaviors they generate are inflexible—the behaviors run to completion.94 The content of the action script is available only to some modules, so other information which may be relevant to behavior is not integrated and the agent cannot interrupt or alter the behavior generated nonconsciously by the action script. Consciousness, however, allows for this information to be broadcast globally and integrated appropriately such that an agent can alter behavior, if necessary. With consciousness, behavior isn’t driven completely by stimuli, but is instead also shaped by how the agent represents the environment and the agent’s beliefs, desires, and values. Consciousness allows for the integration of various content, including new information, environmental cues, etc. (82). This integration of new information allows for agents to be flexible in their behavior rather than to simply run on nonconscious action scripts.95

---

94 Importantly, it is possible for an agent acting without consciousness of some information via an action script to become conscious of the relevant information and thereby interrupt the action script. The claim here is not that an agent who acts on action scripts can never do otherwise, but rather that an agent who acts on an action script without consciousness cannot interrupt or alter the behavior while remaining unconscious of the action script. Levy thinks we act without consciousness of what we are doing via action scripts frequently throughout the day, as when driving a car. But “if a novel stimulus is encountered, we will immediately become conscious of what we are doing…” (2014: 76-77).

95 One might think the consciousness thesis has problematic implications regarding artists and athletes who experience “flow,” or who are “in the zone” while performing. First, it seems such performers, if acting without consciousness, would be acting via action scripts. But such inflexible behavior seems to be precisely the opposite of what we see from such creative minds. Additionally, it seems that these performers would not be responsible for
6.3 Expressing the Real Self

Levy classifies views like RRV as “real self” views of responsibility (2014: 87). Real self views hold that “agents are responsible for actions that express their real selves” (87). Such views, he claims, are actually committed to the consciousness thesis (or should be so committed), whether their proponents realize this commitment or not. The reason, argues Levy, is that consciousness is required for expressing the agent’s true self. Since real self views hold that an agent is responsible only for behaviors or attitudes that express her true self, real self views are committed to the consciousness thesis.

As I explained in the previous section, Levy holds that only consciousness allows for the integration of personal-level content. As a result, behavior that is caused by conscious mental states is more likely to represent an agent’s personal-level attitudes (83). These personal-level attitudes and the response processes to which they give rise can be evaluated for their moral significance only when they can interact with other mental states. In fact, according to Levy, these personal-level attitudes constitute an agent’s evaluative stance, or “the perspective from which we deliberate and evaluate” (90). This perspective requires consistency and coherency:

A set of attitudes must be relatively consistent to constitute a stance: a stance consists of a broad range of mutually supporting attitudes, relatively (though of course never perfectly) consistent and coherent attitudes. An evaluative stance must be relatively unified and coherent, because that’s a condition of agency: if our evaluative stance is not relatively coherent, we would be incapable of pursuing goals or undertaking projects, even short-term ones…. (90-91)

---

their performance if the consciousness thesis is correct. Levy is aware of these concerns, and he draws a distinction between local creativity and innovative creativity. Musicians and athletes experience local creativity, which requires extensive training to acquire various action scripts that are themselves responsive to various stimuli and inputs. Via training, the agent can “break down, combine, and mix” action scripts with other scripts (2014: 119). Once the repertoire of action scripts becomes large enough, the musician’s responses in different musical environments will become increasingly complex and more flexible due to the complexity of the action scripts she has trained in herself and the combination of such action scripts. Nevertheless, Levy claims, these action scripts can run without consciousness. The agent may not be directly responsible for her performance, but she can be derivatively responsible for it because she consciously built up such an extensive database of action scripts. Innovative creativity, on the other hand, requires that one be sensitive to the content of various domains as one combines them in novel ways. Such innovative creativity requires consciousness. For more on this point, see Levy (2014: 119-121).

96 Whether RRV should really be classified as a “real self” view is a matter of debate. Smith herself might resist this classification (see Smith 2008). Nevertheless, I will not dispute this point in what follows.
Levy’s claim is that a nonconscious attitude that gives rise to an agent’s behavior cannot express the agent’s evaluative stance because the agent cannot evaluate that attitude by checking it against other attitudes the agent holds. The behavior might express some attitudes of the agent, perhaps, but these attitudes will be a very narrow range (89). Given that an evaluative stance requires a fuller range of attitudes and commitments, the narrow range of attitudes that might be expressed by such behavior are simply too few to constitute the evaluative stance and thus to represent the agent’s true self.

To illustrate this point, Levy discusses a study done by Eric Uhlmann and Geoffrey Cohen 2005. Subjects were asked to rank candidates based on their suitability for being police chief. Female candidates were consistently ranked as less suitable than comparable male candidates, but subjects insisted that they were appealing to qualifications of the candidates (e.g., formal education vs. “street smarts”) in making a decision and that this decision was objective. Levy argues that even individuals who are “passionately opposed to sexism would not have detected any conflict between the confabulated criterion and their personal-level attitudes” (2014: 95). The sexist attitudes that shaped the subjects’ rankings were not conscious and could not interact with other personal-level beliefs. Additionally, the moral significance of their rankings could not be globally broadcast and assessed. Instead, the confabulated criteria (e.g., that the candidate is formally educated and formal education is more important than street smarts) is broadcast for evaluation (107). Consequently, argues Levy, the selection of male candidates over female candidates could not express the subject’s evaluative stance or true self.

Despite its initial plausibility, I don’t think that Levy’s argument yet shows that expressing one’s real self requires consciousness in the way that Levy claims. For all Levy has said, the following sort of case still seems possible: an agent consciously adopted an attitude in light of other personal-level attitudes and accepted it as part of her evaluative stance. This attitude shaped her evaluative stance, yet over time faded from consciousness, dropping out of the global workspace. If some behavior expresses this nonconscious attitude, the behavior would plausibly nevertheless express the agent’s true self.

This response appeals to a certain understanding of some attitude or state being nonconscious. I am supposing that an attitude can shift from consciousness to nonconsciousness. But this seems to be a safe assumption to make. Personal experience tells us that some attitude
can be conscious and then fade from consciousness. Similarly, an attitude may be nonconscious and rise to the level of consciousness. Levy seems to accept this as well when he writes:

For the purposes of this discussion, I have assumed that the subjects were not aware of the implicit attitude that caused them to confabulate criteria of merit. But there are grounds for questioning that assumption. Agents may be aware of their implicit attitudes. Anyone may perform an Implicit Association Test (IAT), after all, and thereby become conscious of (the content of) some of their implicit attitudes. There is also a growing body of evidence that subjects often know the content of their implicit attitudes. (116)

The mental states with which Levy are concerned seem to be states that could be either conscious or nonconscious. These nonconscious attitudes are not unalterably nonconscious. Other qualities of the mind seem able to call such attitudes to consciousness.

Of course, if an attitude was never conscious to begin with, it would never interact with other personal-level attitudes, and so perhaps could not be part of the agent’s evaluative stance or available in the global workspace. In such a case is the agent not responsible in virtue of the behavior not expressing her real self? I am once again tempted to think we can resist this thought. Even if some attitude was never in the global workspace and so never available for conscious broadcast, the agent’s other attitudes that are in the global workspace might pull this nonconscious attitude into the workspace for evaluation. While Levy claims that being in the global workspace is required for some attitude to express an agent’s real self, perhaps being available to be pulled into the global workspace is sufficient. Consider a case from Travis Timmerman and Sean Clancy: “If an agent were sufficiently concerned about equality, for instance, he would access the attitudes relevant to sexism and change them. The fact that he hasn’t done so might indicate that his other attitudes implicitly ‘approve’ of his sexism, and thus that his real self is (to a degree) sexist” (2015: 111). In this case, despite the fact that the agent’s sexist attitudes are not in the global workspace, the fact that other attitudes in the global workspace do not pull the sexist attitudes into the global workspace for evaluation suggests that the agent’s real self nevertheless may be expressed, despite the relevant attitude not being conscious.
Given these two avenues of response, one might insist that the subjects in Ulhmann and Cohen’s study do express their true selves—or at least that it is possible that they do, for all Levy has shown. In one case, the attitudes that gave rise to the subject’s selection of candidates might have once been conscious and so shaped the subject’s evaluative stance before fading from consciousness. In the other case, the attitudes that gave rise to the subject’s selection of candidates might not have been conscious, but given that other conscious attitudes in the global workspace did not call these attitudes to the workspace, the subject’s choice still expresses the real self. In either case, it seems that the agent might express his real self without consciousness, contrary to the consciousness thesis.

Levy might resist the first avenue of response by protesting that the advocate of RRV is concerned with direct responsibility, and yet such an avenue of response seems to trace responsibility to an attitude’s conscious adoption in a derivative way. This response is unpersuasive, however. RRV can accept that the origin of some attitude and the way that it is formed is relevant to determining the agent’s direct responsibility for that attitude. For instance, if the attitude were formed via some manipulator, the attitude would not reflect the agent’s rational agency. Similarly, if the attitude were formed in such a way that it was never conscious, it might be that the attitude also fails to reflect the agent’s rational agency. Whether this is always the case, however, depends in part upon whether we can be directly responsible for one single attitude.

6.4 The Moral Content and Formation of Attitudes

Levy seems to think that the evaluative stance is the benchmark of agency. But why think that expressing one’s true self, and consequently, direct responsibility, requires the expression of an agent’s evaluative stance? Why, for instance, can one not be directly responsible for one solitary attitude instead? Many nonvolitionists hold that nonconscious, nonvoluntary attitudes (and behaviors) can reveal an agent’s values in virtue of the rational connections between certain judgments and responses. So even if consciousness is required for expressing an evaluative stance, consciousness might not be required for direct responsibility, because expressing an evaluative stance is not required for expressing one’s true self. If this is so, then perhaps an agent can be directly responsible for just one attitude that reflects some judgment.
This seems to be the sort of view nonvolitionists like Smith advocate, and they offer support for the view. It’s uncontroversial that sometimes the attitudes that seem to be reflected in our behavior are not in line with our explicitly endorsed evaluative stance, and some take these attitudes to be more revealing than the behaviors we explicitly endorse. We can test these attitudes using implicit association tests (IATs). One of the most frequently used IATs tests for implicit attitudes toward white and black people. The participant sits in front of a computer with two response keys. There are four different types of stimuli: a black person’s face, a white person’s face, a positive word, or a negative word. As the participant experiences a stimulus, they must categorize the word by pressing either the left or right key. First, the participant categorizes whether the face they see is that of a black person or that of a white person by pressing the appropriate key as indicated on the screen. For instance, the left key may be used for black people’s faces, while the right key may be used for white people’s faces. Then the participant does the same for positive words and negative words. Finally, the categories become disjunctive, so that the participant must classify something as either a black person’s face or a negative word, or as a white person’s face or a positive word. After a few trials, the categories change, such that the participant must classify something as either a black person’s face or a positive word, or as a white person’s face or a negative word. Researchers measure how long it takes the participant to press the relevant key in each scenario. Frequently, they find that it takes participants longer to categorize a black person’s face with a positive word, which suggests that participants have an implicitly negative attitude toward black people.

Many of us explicitly reject racism and sexism, but the results of IATs suggest that most of us have implicit attitudes that shape our behavior in sexist and racist ways (Dasgupta 2004). These attitudes are deeply entrenched in us, and they are difficult to change. And according to McConnell and Leibold 2001, scores on these IATs are better predictors of some racist behavior than are consciously reported attitudes. An agent may report no racist attitudes, and yet he may cross the street to avoid a black person, or use more force when dealing with black individuals. The implicit attitudes and the behavior they trigger need not be conscious. Does the agent reflect a racist attitude for which she is directly responsible?

Angela Smith and T. M. Scanlon both suggest that the answer is affirmative. According to Scanlon, having a pro-attitude toward something is “to see it as reason-giving for us” (2002: 177). Smith says that to have a pro-attitude toward something is to judge it as “good in some
way” (2005: 270). Several of Smith’s examples concern implicit racist or sexist attitudes. Given how Scanlon and Smith understand these implicit attitudes, they seem to function just like conscious attitudes, except they happen to be nonconscious. On their view, there is a connection between whatever the agent desires and what the agent takes to be a reason, and this “makes all our desires constitutive of our practical identity, and makes it appropriate to identify us with each of them” (Levy 2014: 97).

Levy argues, however, that RRV cannot justifiably hold agents directly responsible for their implicit attitudes for several reasons. First, given the way that these implicit attitudes are formed and the way that they function, implicit attitudes lie outside an agent’s rational control. Secondly, implicit attitudes do not have the right sort of robust content to properly ground responsibility. Let’s look at these reasons more closely.

According to Levy, implicit attitudes are quite different from explicit (i.e., conscious) attitudes. Implicit attitudes are gained via associative systems. We see regularities in the environment and come to associate certain items with others. This causes us, for instance, to generate certain feelings of aversion or attraction to the items we associate with punishment and reward, respectively (98). While Smith and Scanlon might insist that finding something motivating entails finding something reasons-giving, Levy disagrees. Because the motivations of implicit attitudes are merely products of associations with punishment and reward, they aren’t really concerned with reasons in the way that Scanlon or Smith claim. And without these reasons undergirding the attitudes, Scanlon and Smith lack the proper foundation to ground responsibility that RRV requires.

In fact, implicit attitudes seem to lie outside the scope of RRV altogether. RRV claims that agents are responsible for certain attitudes and nonvoluntary behaviors because they are judgment sensitive and under our rational control. But implicit attitudes are acquired in ways that bypass rational control altogether, which means that they are not, in fact, judgment-dependent (Levy 2014: 99). Levy claims that these implicit attitudes are a type of *alief*. Tamar Gendler 2011 has coined “alief” to cover “an innate or habitual propensity for a real or apparent stimulus to automatically activate a particular affective and behavioral repertoire, where the

97 Here there may be a difference between Scanlon’s view and Smith’s view. In 5.3.1, I suggested that Smith’s view of rational control requires a capacity. Scanlon, on the other hand, does not seem to require this capacity, as he claims that we can be criticizable for attitudes that are contrary to our settled judgments (1998: 273). If this is an accurate reading of Scanlon, and if rational control requires a capacity, as I assume, then Scanlon does not require that these attitudes be under our rational control.
behavioral propensities to which an alief gives rise may be in tension with those that arise from one’s beliefs”. As aliefs, these implicit attitudes are significantly different from beliefs: “Beliefs change in response to changes in evidence; aliefs change in response to changes in habit” (Gendler 2008: 566). According to Gendler, we might not be able to change aliefs, but merely to bypass them (569). These aliefs don’t respond to our reasons, as anyone who has attempted to alter such implicit attitudes can attest. Even after seeing that one makes certain implicit associations and therefore has implicit attitudes, all one can do is attempt to indirectly influence these attitudes in the same habitual way in which they were formed (Dasgupta 2013). Recalling the reasons why one should not be quicker to associate “woman” with “family” than with “career” has no impact on one’s associations or the behavior that arises from such associations.

It seems like these aliefs don’t involve reasons at all. They are formed as associations, not on the basis of facts that might justify them. This is especially concerning for an advocate of RRV who understands responsibility as answerability. On this view, agents are answerable for some action or attitude when they can appropriately be asked for their justificatory reasons for that action or attitude. But as Levy notes, “[W]hen an action is caused by an implicit attitude that is acquired and persists for no (justifying) reason at all, the agent can’t cite justificatory reasons for the action (though they may, wrongly, take themselves to be able to do so…)” (2014: 100). Given the way implicit attitudes are formed and the way they function, there seems sufficient reason to think that RRV cannot account for our responsibility for these implicit attitudes directly. Not only are such attitudes not formed for reasons, but they also seem to lie outside our rational control.

Levy’s second reason for excluding implicit attitudes from the scope of direct responsibility is related to the attitudes themselves. He argues that implicit attitudes lack the robust content required for moral responsibility. The content of an implicit attitude is “the associations it activates and the related content it primes” (102). Even though these implicit attitudes cause racist and sexist behavior, all that is expressed in the implicit attitude is a statistical association between two items, such as “woman” and “family,” or “male” and “police chief”. But, Levy contends, there is nothing here that is morally condemnable: “the fact that I associate $X$ and $Y$, nonconsciously, is no basis for holding me morally responsible” (102).

One might contend that the association itself is, in fact, morally condemnable. After all, the association itself is either racist or an indication that the agent is racist. Recall, however, that
these associations are formed as a result of reward and punishment. One’s association of black people and violence is not formed on the belief that black people are violent, for instance. For this reason, Levy argues:

[The association] does not license the inference that would have to be true for it to be the case that the attitude was racist: it does not license inferences to conclusions like the agent believes (nonconsciously) that [black people] are inferior, or even to the conclusion that the agent believes (nonconsciously) that [black people] are violent. The latter attributes to the agent a sufficient number of the dispositions constitutive of the stereotype associated with such a belief, but the agent possesses such dispositions (again, even nonconsciously) only in virtue of having a conscious belief in claims like this. There is a temptation to think that the attitude in virtue of which the agent responds as he does is a nonconscious belief, but this is a temptation that should be resisted. (105-106)\(^98\)

Levy’s claim is that these associations are not morally robust enough to be morally condemnable on their own. Nor are these associations morally robust enough to warrant inferences to an agent’s morally condemnable beliefs. Implicit attitudes are too thin to be items for which we are directly responsible, and they lie outside our rational control. As a result, they are not items for which we can be directly responsible, even on RRV.\(^99\)

It seems that Levy wants to apply his remarks about implicit attitudes to any attitude that is nonconscious:

[T]he temptation to think we can ground (direct) moral responsibility in expression of nonconscious attitudes attributes to these attitudes features that are the province of our conscious attitudes alone. Our nonconscious attitudes are too thin to plausibly be regarded as constituting reasons for us, even nonconscious reasons. Moreover, their content is too distant from the kinds of content that could plausibly ground moral responsibility. They are mere associations, not attitudes in virtue of which we can

\(^{98}\) I have changed “blacks” to “black people” in the above quote.

\(^{99}\) Of course, agents who are aware that they have implicit attitudes or make these implicit associations can be held derivatively responsible for failing to attempt to take certain steps to correct these attitudes. We can hold them responsible for their efforts. If Levy is correct, however, we cannot hold them directly responsible for these implicit attitudes themselves.
reasonably be blamed. This is true, even though these attitudes are often the cause of conscious attitudes that do seem well suited to play the role of underwriting moral responsibility. (107-108)

This generalization is too quick, however. It is unclear from the passage above why Levy thinks that we cannot ground direct moral responsibility in expression of nonconscious attitudes unless he thinks that every nonconscious attitude is an alief (i.e., an implicit attitude that was formed via associations and which lies outside of our rational control). But this does not seem to be the case, or at least Levy has not given reason to think otherwise. Levy’s arguments at most show that we cannot be directly responsible for aliefs. Attitudes that are within our rational control but which happen to be nonconscious have not yet been ruled out.

Perhaps a nonconscious attitude must be an alief unless it was formed consciously and assessed and held for reasons. But even if this is the case, RRV is not in as dire straights as Levy might suggest. Provided that it is possible for an attitude to be consciously adopted as part of one’s evaluative stance and then relegated to nonconsciousness, there can be nonconscious attitudes that are not aliefs. Such an attitude could be formed via rational processes, and it could have robust moral content that is more than mere statistical association, yet fail to be conscious in the agent’s mind. If this is right, then Levy has not yet shown that the consciousness thesis is true.

6.5 The Mildly Limited Scope of RRV

Confident in the truth of the consciousness thesis, Levy writes, “I think there is less at stake in the debate between the two rival accounts of moral responsibility mentioned here than the proponents of these views believe: they excuse the same agents under the same conditions” (106). RRV’s attraction was that it apparently offered an explanation of responsibility that required minimal revisions to our ordinary judgments. Yet if the consciousness thesis is true, then it seems that we might be forced to revise some of these judgments after all.

I have given reasons to think that the consciousness thesis is false. Agents can be responsible for attitudes (or behaviors that reflect some attitude), even if those attitudes (and
their moral significance) are not conscious. But Levy provides powerful reasons for thinking that agents cannot be directly responsible for implicit attitudes that are aliefs. While Levy’s argument concerns consciousness rather than voluntariness, this argument has important upshots for RRV. Given RRV’s criteria and empirical data regarding implicit attitudes, it seems that implicit attitudes lie outside of our rational control and consequently cannot be within the scope of direct responsibility. While advocates of RRV like Smith may have thought the view captured all implicit attitudes within direct responsibility, this was a mistake. Agents are plausibly not directly responsible for their aliefs, as such attitudes are not judgment sensitive. Of course, if these kinds of implicit attitudes are not within an agent’s rational control, this is the right result for RRV, and these attitudes should not be items for which agents are directly responsible. Levy’s argument creates no large problem for RRV, but merely limits the scope of those items for which agents are directly responsible.

Importantly, however, the scope is not limited to the extent that Levy thinks. If not all nonconscious attitudes are nonrational in the way many implicit attitudes are, then RRV can still account for direct responsibility for a variety of nonconscious attitudes—even if those attitudes and their moral significance fail to be conscious. Nonconscious attitudes can still reflect some judgment, and possibly even an agent’s real self. Despite Levy’s objections, RRV remains a plausible account of moral responsibility that can account for a variety of our ordinary responsibility ascriptions and practices.

---

100 Levy states the consciousness thesis as pertaining to facts that give one’s actions moral significance, but his discussion often simply focuses on the consciousness of an attitude rather than its moral significance. It is unclear how Levy understands these two to be related. I take it that in being conscious of the moral significance of one’s actions, one is conscious of certain attitudes. Thus one must be conscious of certain attitudes in order to be conscious of the moral significance of one’s actions or attitudes.
CHAPTER 7

THE RATIONAL CONTROL VIEW

7.1 Introduction

It is useful now to pause and take stock. I began in chapter 2 by offering VT2, which I take to be a clearer version of the volitionist thesis than that which is currently discussed in the literature. Because VT2 places many omissions outside the scope of direct responsibility, I examined amended versions of VT2, one of which is broader than the other. I argued that the broad thesis, VT3, lacks a sufficient justifying account of why agents are directly responsible for their omissions, and attempts to remedy this by adding a quality of will condition bring VT3 close to nonvolitionism. VT4 does make room for quality of will considerations and provides an account that can justify responsibility ascriptions and practices, but does so at a price. It does not seem that the same reasons that justify direct responsibility for intentional actions can justify expanding direct responsibility to include intentional omissions as well. As a result, neither VT3 nor VT4 seem sufficiently justified or motivated as plausible necessary conditions on direct responsibility.

These considerations do not mean that volitionism should be abandoned. One may accept VT2 and its restricted scope of direct responsibility, but account for our responsibility for a variety of other items derivatively. Agents may be derivatively responsible for consequences, omissions, negligence, and attitudes, even if agents are not directly responsible for these items. This may be a difficult line to press, however, as I argued in chapter 4 that it is not clear that volitionists can account for many of these items derivatively.

In light of these shortcomings, I suggested that there is sufficient reason to entertain nonvolitionism as a contending theory of moral responsibility. I turned in chapter 5 to what I take to be the most plausible and developed nonvolitionist view currently on offer: Angela Smith’s Rational Relations View (RRV). Although I have argued that a view like RRV can withstand many of the common objections raised against it, I also allowed that at times a proponent of such a view must make some concessions or modify some of the claims of the
view. As a result, the nonvolitionist view I contend may stand as a plausible alternative to volitionism remains as of yet somewhat amorphous. While I cannot offer a full defense of this view here, in this final chapter, I will offer some formal conditions on direct responsibility, taking my cue from RRV. I call the view I propose the Rational Control View (RCV), which is deeply influenced by RRV, but nevertheless stands apart from the view in its explicit commitments and applications. While I cannot fully defend RCV here, I hope to elucidate and motivate the view, explaining its commitments and defending it from some objections. Given its ability to better account for our everyday responsibility ascriptions and practices, I hold that RCV is a strong contender for a plausible theory of moral responsibility. In outlining and defending RCV’s commitments here, I hope to show that we have good reason to take RCV seriously, and possibly even to accept RCV over its volitionist alternatives.

7.2 The Rational Control View

RRV focuses on responsibility for attitudes via an agent’s rational control. Adopting and formalizing this basic idea, the defining thesis of RCV is

\[ RCT: \text{An agent is directly responsible only for what is under her direct rational control.} \]

RCT restricts direct responsibility to rational control rather than intentional action or voluntary control. This sets RCV apart from volitionist views. But RCT requires an account of direct rational control, which I offer here:

\[ RCd: \text{An item, } A, \text{ is under } S' \text{ ‘s direct rational control just in case (i) } A \text{ is some judgment of } S' \text{ ‘s or at least partially constituted by some judgment of } S' \text{ ‘s, and (ii) } A \text{ is normatively connected to } S' \text{ ‘s judgments and values such that } A \text{ is open to revision via rational reflection.} \]

\[ It \text{ is unclear to me whether RCV is nothing more than RRV with the revisions that I think are appropriate to make the view plausible, or whether it stands on its own as a separate view. I nevertheless adopt a different name for the view I propose to clarify Angela Smith’s commitments as opposed to my own. Regardless, much of my thinking here is indebted to Smith.} \]
I explained in 5.3.1 how I understand revision via rational reflection, but I offer two conditionals to elucidate the idea here. One is for attitudes (represented by the variable $A$), and one is for evaluative judgments (represented by the variable $V$):

**RRa:** $A$ is open to revision via rational reflection if it is the case that, if $S$ were to continue to exist and to cease to have some judgment, $V$, then, if left to her own devices, either $A$ would be altered or else $S$ would cease to have $A$.

**RRv:** $V$ is open to revision via rational reflection if it is the case that $S$ is capable of judging that some reason $R$ is sufficient to give up $V$, and if $S$ so judged and did not hold any other judgments that conflicted with $R$, then $S$ would cease to have $V$ because of $R$.

Although I have discussed the content of RCd, RRa, and RRv in previous chapters, it is useful to offer a few further remarks on them here.

First, note that RCT focuses on *direct* responsibility. This is part of what sets up RCV as an alternative to volitionist views. Next, RCT appeals to *direct* rational control. The first conjunct of RCd is what makes rational control direct. Our actions and omissions are normatively connected to our judgments and values, but our actions and omissions are not *constituted* by some judgment we hold. I followed Angela Smith in appealing to this point in 5.3.3. Importantly, however, given this account of RCd, I am committed to the claim that agents are directly responsible only for their evaluative judgments and those attitudes that are at least partially constituted by judgments. Our judgments and attitudes merely cause our actions and omissions rather than constituting them in a more intimate way. While Smith seems to want to resist this claim in explaining RRV, this seems to be the right move to make, given the nature of direct responsibility. When an agent is responsible for some item in virtue of being responsible for some prior item, the agent is derivatively responsible. And it seems plausible that one is not responsible for some action or omission if one is not responsible for the attitude that causes the

---

102 These are modified versions of Smith’s conditions, as presented in Miller 2014. As I argued in 5.3.1, RRa and RRv are meant to show that a capacity is required. (I have dropped the “in principle” for this reason.) Admittedly, counterfactuals do not seem to be extensionally equivalent to capacities. For instance, capacities admit of failure, but counterfactuals seem to leave no room for this failure. Despite the differences in capacities and counterfactual claims, I use the counterfactuals of RRa and RRv as a *rough* expression of the capacity I described in 5.3.1. I have made certain modifications to avoid a few counterexamples that pose problems for the conditions Smith offers. Smith may be amenable to these modifications herself, but I do not apply them to her original conditions in 5.3.1.
action or omission. If our responsibility for our behaviors depends on our responsibility for the attitudes that are reflected in those behaviors in this way, our responsibility for our behaviors is derivative rather than direct. Of course, the trace from the attitude to the behavior is very short, but the responsibility is derivative nonetheless. Finally, as I explained in 5.3.1, RRa and RRv require that one have the capacity to revise one’s attitudes or judgments. This capacity is required for rational control.

RCT is one necessary condition on direct responsibility, but it is meant to stand as an alternative to VT2, VT3, and VT4. While it parallels these volitionist theses in its structure, one might wonder if more can be send to sharpen the contours of RCV. Offering careful and accurate sufficient conditions on moral responsibility is too large a project for me to undertake here, but I will offer a sketch of some plausible further necessary conditions. These conditions, or something like them, might be jointly sufficient for responsibility.

7.2.1 History and Manipulation

First, there are good reasons to think that the history of some attitude also matters for responsibility. If an attitude is not formed in the right way, the attitude cannot count as the agent’s, or as one that reflects the agent’s rational processes.

Angela Smith does not seem to have a robust historical condition for RRV. She does want to rule out cases in which an agent’s judgments or values are implanted, and so “not the agent’s own.” She writes:

Implanted attitudes, by contrast, tell us nothing about the agent’s rational or evaluative dispositions. I see no other way of giving content to the expression ‘the agent’s own’ here, however, except in a way which makes reference to the very network of beliefs and attitudes which I am suggesting ground our attribution of responsibility. […] A reasonable account of the conditions of responsibility should preserve our sense of the rational interrelations among our attitudes.... (2005: 262)

According to Smith, RRV rules out responsibility for manipulated attitudes via some sort of coherence condition:
CC: $S$ is responsible for some attitude, $A$, only if $A$ coheres to some extent with the agent’s network of beliefs and attitudes.

One interpretation of CC might be friendly to Levy’s arguments in chapter 6. Perhaps CC should be read as requiring that an attitude be a part of an agent’s evaluative stance in order for an agent to be responsible for the attitude. If consciousness is required for an attitude to be adopted as part of the evaluative stance, as Levy contends, then perhaps consciousness is required for CC to be satisfied. CC may be read in a weaker way, however, such that the attitude must simply cohere with the evaluative stance without being adopted into the stance. On this reading, consciousness is not required.

I suspect Smith would favor the second reading over the first. She seems to want to hold that an agent can be directly responsible for some attitude even if that attitude is out of character and conflicts with explicitly held beliefs that the agent has. For instance, Smith holds that we are responsible for our implicit attitudes. Yet as we have seen, an agent’s implicit attitudes might not cohere with her network of beliefs and attitudes.

I will not attempt to elucidate CC further, as I do not think that CC can, on its own, do all the work it purports to do. Daniel Miller has recently argued that CC cannot rule out all problematic cases of manipulation, such as when an agent is manipulated to hold a coherent network of beliefs and judgments. We do not think that such agents are plausibly responsible for such judgments or the responses arising from them (2014: 474 ff.). To rule out this possibility, Miller has proposed a different historical condition on RRV, which I have adapted here:

\[ HC: \text{An agent } S \text{ is directly responsible for an attitude } A \text{ at } t_1 \text{ only if it is not the case that (a) } A \text{ was acquired at some earlier time } t_0 \text{ in a way that bypassed } S \text{’s rational capacities completely and (b) } A \text{ is unsheddable from } t_0 \text{ (at least) up to and including } t_1.^{103} \]

An unsheddable attitude is one that an agent is practically unable to eradicate or significantly attenuate.\(^{104}\) While an attitude or judgment must be under $S$’s direct rational control, it is not

---

enough, on RCV, that the attitude or judgment currently be under S’s rational control. It must not have been gained in a way that bypasses that rational control.

While Miller uses HC to rule out manipulation cases for RRV, HC also serves to rule out direct responsibility for implicit attitudes that were not formed via rational processes. If implicit attitudes are simply aliefs, and aliefs are unsheddable and are acquired in ways that bypass rational capacities, as Gendler and Levy argue, then they do not satisfy HC. In this way, HC is compatible with Levy’s claim that agents cannot be directly responsible for these kinds of implicit attitudes.

7.2.2 Ignorance, Awareness, and Expressing a Judgment

I argued in chapter 6 that the consciousness thesis is false. One need not be aware of the morally relevant features of one’s action to be morally responsible for that action. This raises important questions for RCV about ignorance and moral responsibility. Following Matthew Talbert, we can distinguish between two types of ignorance: moral ignorance and circumstantial ignorance. Moral ignorance is ignorance of moral facts or principles. Circumstantial ignorance is ignorance of one’s circumstances. Many nonvolitionist views do not hold that moral ignorance excuses an agent from responsibility. This is not always the case for circumstantial ignorance.

Circumstantial ignorance may sometimes be relevant to whether an agent is responsible, as it can determine what judgment the agent expresses in her behavior. Matthew Talbert writes:

[J]udgments about a circumstantially ignorant wrongdoer’s moral responsibility should track the plausibility of associating her behavior with interpersonally significant evaluative judgments—particularly judgments about the normative status of the needs and interests of those affected by the wrongdoer’s actions and omissions. (2013: 235)

If Simone knows that her behavior will harm Matthias, then we can reasonably infer that her behavior expresses an objectionable judgment that Matthias’s harm is not a decisive reason to refrain from the behavior. In order to reasonably attribute this judgment to Simone, however, we must assume that Simone has some circumstantial awareness of her situation: Matthias has

---

105 Notice that Talbert is focused on behaviors here. On RCV, agents are derivatively responsible for behaviors in virtue of being responsible for the judgments these behaviors express. As in chapter 6, I will often focus on behaviors to engage with the writers under discussion.
interests and can be harmed, and engaging in some behavior will have the result that Matthias is harmed, etc. If Simone is unaware of the fact that her behavior will harm Matthias, it seems implausible to attribute to her any sort of objectionable judgment, such as “Matthias’s interests don’t matter.” In order to express a judgment that someone’s interests don’t matter in one’s behavior, one must be aware that in these circumstances, one’s behavior discounts or affects someone’s interests in a negative way.

Suppose that Simone is aware of these circumstances, but is morally ignorant of the fact that the fact that her behavior will harm Matthias is a decisive reason not to engage in that behavior. Is Simone’s moral ignorance excusing? It seems not, as Simone’s behavior expresses an objectionable judgment that Matthias’s interests do not matter. Importantly, her behavior expresses this judgment even if she is not culpable for being ignorant of this moral principle. Simone’s moral ignorance in this case is “tied up with a perspective that regards as unobjectionable the very thing that [her] victims take to make [her] behavior objectionable” (Talbert 2013: 236). Given that Simone’s behavior expresses an objectionable attitude that is under her rational control and was not formed by bypassing her rational capacities, there seems to be no good reason why she is not responsible for this behavior or for the attitude that it expresses. Her behavior is, as Talbert might say, “knowing, deliberate, and dismissive of the interests of those [her] actions affect” (238). One might protest that it would be unfair to blame Simone if she is nonculpably ignorant of some moral principle, but I have already addressed these concerns in 5.3.2 when appealing to Hieronymi’s argument. As long as the agent expresses an objectionable judgment that was under her rational control and was not formed in the wrong way, she may be responsible for the judgment.

Given the claims that moral ignorance does not excuse but circumstantial ignorance may excuse, I suspect that the former is more controversial. It is this first claim that Neil Levy aims to attack when he offers the following case:

Suppose that there is a kind of harm that is objectively morally relevant, but of which we are ignorant. Suppose, for instance, that plants can be harmed, and that this harm is a moral reason against killing or treading on them. In that case, many of us are causally responsible for a great many moral harms. Are we morally responsible for them? Do we flout a moral requirement, and challenge plants’ standing as objects to which some moral
consideration is owed? No to all these questions: If we do not grasp the moral requirement, and this ignorance is not culpable, we do nothing blameworthy. (2005: 9)

Levy’s case is intended to show that our moral ignorance excuses us from blameworthiness in this case, so moral ignorance can, at least sometimes, excuse. But Talbert convincingly argues that Levy’s case is not one of moral ignorance, but rather of circumstantial ignorance:

Levy’s example fails to illustrate the exculpatory power of moral ignorance because what explains why we are not blameworthy in the example is that we lack crucial information about how walking on plants affects them. Since I do not know that plants can be harmed, my stepping on one does not express a denial of the significance of its being harmed. So, on the view I advocate, stepping on a plant does not express a judgment that could properly ground blame. (2013: 243)

Levy’s case is a close parallel of the case of Simone and Matthias. Talbert cannot express an objectionable judgment, such as “The interests of plants don’t matter,” because Talbert isn’t aware that plants have interests to matter in the first place. Without this circumstantial awareness of the fact that stepping on plants harms them and violates their interests, Talbert’s behavior cannot be said to express any judgment that the interests of the plants don’t matter.

Of course, there is some kind of morally relevant ignorance in Levy’s case, but this ignorance is rooted in circumstantial ignorance. It is not ignorance of some moral principle, as Talbert puts it. Talbert writes, “I presumably know that I generally have moral reasons not to cause pain; I just don’t know that stepping on plants causes them pain” (243). But what about a case in which Talbert is not circumstantially ignorant, but is merely morally ignorant? In this sort of case, Talbert would know that stepping on a plant causes it pain, but he believes (nonculpably, we can suppose) that this pain doesn’t matter. With Talbert, I would hold that in this case, Talbert is blameworthy: “If I knowingly (and unjustifiably) cause a plant severe pain, then I am open to resentment and indignation because my action dismisses the normative significance of the plant’s pain, and this is something to which the plant (or at least one who is concerned for its welfare) could reasonably object” (243-244).
Compare Levy’s plant case with Gideon Rosen’s Hittite lord in the ancient Near East who buys and sells human beings. The lord is aware that anyone could become a slave through bad luck, and presumably he is aware that human beings can suffer and feel pain. Yet he is ignorant of some moral principle, it seems, as “the legitimacy of chattel slavery was simply taken for granted” (2002: 64). Though Rosen judges that the Hittite lord is not blameworthy for selling slaves because he is nonculpably ignorant of the moral principle that forbids this, according to RCV the lord is blameworthy in virtue of the judgment that his behavior expresses. Because he is aware that human beings experience pain and have interests, and because he is aware that the slaves he buys and sells are human beings, he expresses a judgment that the interests of the human beings do not matter. And this is a judgment that is objectionable. If the Hittite lord did not realize that the slaves he was buying or selling could feel pain, however, his circumstantial ignorance may prevent him from expressing an objectionable judgment for which he is blameworthy.

It is important to note that even though RCV holds that those who are nonculpably morally ignorant are nevertheless responsible, this does not yet entail that certain people ought to blame such agents for their behavior. If everyone in the ancient Near East was nonculpably ignorant of the moral principle forbidding slavery, then presumably none of them had the standing to blame the Hittite lord for his behavior on pain of hypocrisy. The Hittite lord was blameworthy, but none of his contemporaries could blame him in virtue of their lack of moral standing. Some possible agent who did have the standing could blame the Hittite lord—God, perhaps—so the Hittite lord is still blameworthy despite his nonculpable moral ignorance.

I claimed above that circumstantial ignorance may excuse, but it doesn’t seem that it always does. For instance, there is an important difference between the circumstantial ignorance in Levy’s case or Simone’s case and the circumstantial ignorance that might be involved in some cases of negligence. Consider a doctor who fails to look at a patient’s chart to verify that the patient has no allergies to a certain medication, but who is aware that some medicines can cause patients with allergies harm. As a result, the doctor is ignorant of the fact that the medicine she prescribes is harmful to the patient. Yet she nevertheless expresses something morally significant.

---

106 For an argument that hypocrisy can undermine one’s standing to blame, see Fritz and Miller n.d.
in her behavior. In this case, her behavior seems to reflect a lack of sufficient concern or care for her patient. If she cared more about her patient and her patient’s welfare, she would have reviewed the chart before administering the medication. The doctor may be ignorant of some features of her circumstances, but this ignorance does not keep her from reflecting this lack of care. She understands that patients have interests and can be harmed and that some medications can harm patients if they have certain allergies. This knowledge is sufficient for the doctor to reflect an objectionable lack of judgment in this case.

7.2.3 Comparing RCV with Alternatives

While there is still much work to be done in explaining when an objectionable judgment is expressed in some behavior and the role of ignorance in this connection, these remarks can help to clarify some of RCV’s commitments—including its commitment to the claim that moral ignorance does not preclude responsibility. My goal here is simply to outline what I take to be plausible contours of RCV that can help to distinguish it from RRV as well as to set it up as a plausible alternative to volitionist views. In that regard, there are some important features to note about RCV as I have described it. First, the view is unified. I criticized both the traditional understanding of “voluntary control” and VT4 as being problematically ad hoc. But RCT appeals only to one type of control: rational control. It makes no appeal to tenuous or unclear notions of voluntary control or the voluntary.

Rational control, on RCV, is the fundamental underlying grounding that justifies responsibility ascriptions. The evaluative attitudes we hold are the foundation of our

---

107 It is not clear that negligence always involves an objectionable judgment (or lack thereof), and so it may be that not all cases of negligence are ones for which the agent is blameworthy. Perhaps Matt King’s example of Negligent Nate (see chapter 4) or George Sher’s Hot Dog (see chapter 5) are cases of this sort. Precisely when an objectionable judgment is reflected in behavior is an important question that nonvolitionists must explore much more deeply to clarify their view.

108 Sometimes agents express an objectionable judgment in their behavior, and sometimes they reflect an objectionable lack of some judgment in their behavior. While the doctor may express some objectionable judgment, such as “The potential harm to the patient is not worth reviewing the chart,” or “The patient’s interests are not sufficient reason to review the chart,” this seems more akin to recklessness than negligence. Negligence often involves a lack of some attitude that one ought to have. What we find objectionable in such cases is that agents fail to have certain attitudes that they should, and this failure is itself what is objectionable.

109 If the doctor was unaware that allergies could interact with medications in a way that would harm her patient, then this case would be akin to Levy’s plant case, and the doctor would not plausibly be blameworthy for administering the dangerous medication. This would be circumstantial ignorance that would preclude us from making the inference to her objectionable lack of judgment (or objectionable judgment).
Sometimes these attitudes cause and are indirectly reflected in intentional actions and omissions. Sometimes, however, they cause and are reflected in nonintentional actions and omissions. In either case, it is rational control that remains the important feature for responsibility ascriptions. And it is because of its focus on rational control that RCV can account for responsibility for both intentional and nonintentional actions and omissions. In virtue of the above features, RCV can accommodate a widely held distinction between some item being bad and some item being blameworthy. Neil Levy presses this point against Smith’s RRV: “We can hold that assessment of attitudes is simply that: the attribution of qualities to the agent that are good or bad, admirable or repugnant. After all, prima facie there ought to be conceptual space for such assessment. We ought to be able to say that something is bad without saying that it is blameworthy” (2005: 5). While RCV doesn’t hold that the assessment of attitudes is where we draw the distinction between something being bad and something being blameworthy, it allows that an agent may simply be assessed as a bad agent in virtue of some feature without being blameworthy for that feature. This is an important and intuitive distinction that any reasonable view of responsibility ought to make room for. It is a virtue of RCV that it can account for this distinction.

While agents may be blameworthy for their attitudes, for instance, they are directly blameworthy for some objectionable attitude only if they satisfy the conditions above (i.e., RCT and HC). As I have argued, implicit biases likely do not satisfy these conditions, and so agents will fail to be directly blameworthy for objectionable implicit biases. Nevertheless, we can admit that such biases are bad and that it would be better if agents didn’t hold them. Smith, however, acknowledges that she cannot accept the distinction between bad and blameworthy agents in many instances (2008: 387-390). Insofar as Smith insists that implicit biases are held for reasons and under our rational control, she cannot leave room for a distinction between a bad attitude and a blameworthy attitude.

I take it that Hieronymi (2014: 9-16) makes a similar point much more carefully and forcefully than I do here when she claims that answerability fundamentally explains how we are responsible for our intentional actions. See also Hieronymi 2008.

Holly Smith 2011 argues for a condition on direct blameworthiness for the nonvoluntary. However, her condition is meant to be joined alongside a condition on direct blameworthiness for the voluntary. In appealing to two separate conditions, one for the voluntary and one for the nonvoluntary, Smith abandons a unified view. It is an open question what unifies these two conditions on blameworthiness. RCV avoids this problem by pointing to the underlying feature that unites responsibility ascriptions: rational control.

See Smith n.d. for more details.
To summarize, while RCV takes much from RRV, it differs in at least three important ways. First, it includes the historical condition HC. Second, it denies that agents are directly responsible for implicit biases because such attitudes are really aliefs and are not under an agent’s rational control. Finally, in virtue of these features, it accommodates a distinction between some attitude being bad and an agent being blameworthy for some attitude. I take each of these differences to be virtues that RCV has over RRV, and thus reasons to favor RCV over RRV.

7.3 Objections and Replies

In chapters 5 and 6 I defended RRV from several objections. Many of these objections, as well as my replies, apply to RCV as well. Below, I examine a handful of lingering objections that can help elucidate and strengthen RCV.

7.3.1 Evaluative Stance vs. One Attitude

According to RCV, one is responsible for some attitude or behavior when that attitude or behavior reflects some evaluative judgment that was under one’s rational control. But one might complain that reflecting one judgment is too low of a bar; one’s attitude or behavior must be reflective of one’s evaluative stance. It is the agent we hold responsible, and the agent should be associated with her evaluative stance, not one attitude she holds. As Levy puts the point, “We cannot praise or blame, punish or reward an attitude or a small set of attitudes; rather it is the agent who is the target of our reactive attitudes” (2011: 257).

Holly Smith offers a case to illustrate this point:

_Hypnotized:_ Clara, a student, has a desire to wound her peer, Bonnie, with a hurtful remark. But Clara also desires to maintain a good reputation and to impress her boyfriend, and she knows saying the hurtful remark to Bonnie would jeopardize these other desires. During her psychology class, Clara is hypnotized such that some of her desires, like the desire to impress her boyfriend and to maintain a good reputation, are

\[113\] Here I intentionally blur the distinction between direct and derivative responsibility. RCV holds agents directly responsible only for attitudes, so agents will be only derivatively responsible for their behaviors that reflect these attitudes.
inhibited. As a result, Clara says the hurtful remark to Bonnie from her desire to wound her. (adapted from Smith 2011: 133-134)

According to Smith, Clara isn’t blameworthy. But this is not because she is hypnotized. After all, argues Smith, the desire to wound Bonnie is not implanted by a hypnotist, but is genuinely one of Clara’s desires. Instead, Clara isn’t blameworthy because,

although her choice did arise from her own desire, and does reflect her entire evaluational response at that moment to her options, the choice did not arise from anything like a reasonably full configuration of the motives that she actually has and that would normally bear on such a decision. Her choice reflects *part* of Clara’s psychology, but not *enough* of her psychology to warrant a judgment of blameworthiness. To blame someone is to negatively evaluate that person’s whole motivational structure insofar as it bears on the choice, and Clara’s choice does not reflect anything like her full personality. (134)

The idea is that we hold an agent responsible in virtue of her moral personality, but one attitude doesn’t reflect enough of an agent’s moral personality for this responsibility judgment to be justified.

Of course, one might be wary of hypnotism cases, as it might be the case that the agent is not responsible precisely because she was hypnotized, and not because the attitude fails to reflect her full moral personality. The hypnotist prevents Clara from responding as she would normally, and one might think that this diminishes Clara’s responsibility. There are several other examples of cases in which agents “are often caused to behave in certain ways by environmental stimuli while remaining completely unaware of the role these stimuli play in causing their behavior” (134). There is a growing body of evidence that situational factors and influences affect behavior without the agent’s knowledge.

In one case, participants who were presented with synonyms for rudeness on a test were subsequently rude in a situation staged soon afterward. This is in contrast to participants who

---

114 It doesn’t seem that RCV has the resources to clearly resist the conclusion that Clara is responsible (and blameworthy) in this particular case, however, as the hypnotist does not violate HC. Clara’s objectionable attitude toward Bonnie, for all we are told, might be formed for reasons and might be under her rational control. The remark would therefore reflect an objectionable attitude on her part. Nevertheless, I don’t take this to be a particularly disconcerting problem for RCV, as we can appeal to degrees of blameworthiness here. I expound on this idea below.
were presented with synonyms for politeness, who behaved politely in the staged situation (Bargh 2005: 37). In another case, when an elderly stereotype was subtly activated in participants’ minds, they walked more slowly (Chartrand and Bargh 1999: 893-910). Participants who are asked about their attitudes towards gay men respond with more negative attitudes if they are also near a foul odor (Liberman and Pizarro 2010).¹¹⁵

Many more cases have been offered. Much of our behavior is not consciously guided and seems to be influenced by aspects of our situation and environment. As Dana Nelkin argues, many of us act for reasons that we do not realize, and we seem to be stuck doing so (2005: 199). While we may not be hypnotized, we frequently act in ways that reflect only one attitude or a small subset of attitudes, and part of the reason for our acting in these ways is due to environmental factors of which we are unaware. The problem, according to Smith, is that many of these environmental stimuli trigger automatic processes. This automatic behavior that issues cannot reflect the agent’s full set of evaluative attitudes toward the behavior, as there is no weighing of desires or motivations. Such automatic behaviors sound much like Levy’s action scripts, which run without conscious awareness. This automatic behavior, according to Smith, “is more or less isolated from the agent’s other motivations, and so does not adequately reflect the agent’s full psychology, or full evaluation of the act to be performed” (2011: 137). Yet it is the agent we are holding responsible, so it is unfair or unreasonable to blame an agent for some automatic process that does not reflect enough of her full moral personality.

Smith does not require that the agent’s entire set of pertinent desires play a role in some process for the agent to be responsible for what arises from that process (138). While it is difficult to draw some line, Smith is concerned here with a sufficiently complete set of desires and aversions, which may be significantly less than an agent’s full set of desires and aversions. The general point, for Smith, is that we cannot justifiably hold agents responsible for behavior that reflects only one attitude or for behavior that does not reflect enough of her evaluative stance. To do so would be unfair given that much of this behavior is automatic and without consciousness.

¹¹⁵ There are concerns about whether these studies can be successfully replicated (see Doris 2015: ch. 3). While I think this is a genuine concern, I take it that the general volume and type of evidence is at least reason to think that situational factors can and do influence our behavior in significant ways, and that we are unaware of many of these factors as well as their influence.
In response, it is important to reiterate that if my responses to Levy in chapter 6 were correct, some nonvoluntary, nonconscious behavior may still reflect an agent’s evaluative stance, provided that the attitude reflected in the behavior was formed consciously and shaped one’s evaluative stance before fading to nonconsciousness. If that attitude is in line with the agent’s evaluative stance and is strong enough, some automatic behavior may yet reflect enough of the agent’s moral personality to count as something for which the agent is responsible. Smith admits as much, but thinks such cases will be rare. How frequent these cases are is a question for empirical investigation, but one need not insist that agents are only responsible when some behavior reflects their evaluative stance. Much of our behavior is not as reflective as Neil Levy or Holly Smith require for responsibility. If this is the case, and if this reflection is required for moral responsibility, we are responsible for much less than we originally thought.

There is good reason to resist this skeptical line of reasoning. While one attitude may not fully represent an agent, it represents enough of an agent to hold that agent responsible. There is a disagreement here about thresholds. Where is the threshold of responsibility? For RRV and RCV, the threshold is one attitude or judgment. But it is important to keep in mind that these attitudes and judgments are often connected and importantly related. Our judgments are not made in isolation of each other, but rather influence each other and influence the ways in which we view the world. For Levy and Holly Smith, however, the threshold for responsibility is much higher: the evaluative stance. I suspect that ultimately, little can be said to sway those on one side to the other. But there is something that might make the idea that we are responsible for one attitude, as well as for behavior that reflects that attitude, more palatable to theorists like Levy and Smith.

Blameworthiness (and praiseworthiness) come in degrees. Nonvolitionists can and should make use of this notion of degrees. Suppose an agent has an objectionable attitude, and that attitude is manifested in some nonconscious, automatic behavior. If that attitude was not formed in such a way that it violates HC and is under the agent’s rational control, then the agent is responsible for that attitude and for the behavior that reflects it. But it is a separate question how blameworthy the agent is for the attitude. Agents seem to deserve more blame if they knowingly and willingly engage in objectionable behavior.

---

116 Holly Smith points out this omission in the literature: “Typically Attributionists don’t address the question of whether the culpably ignorant agent is as culpable as the witting agent who performs the same act” (2011: 119).
What explains this difference? One plausible explanation is that agents reflect a more objectionable attitude when they knowingly and willingly behave in objectionable ways. When some agent intentionally and consciously stomps on my hand “with a malevolent wish to injure me” (to borrow Strawson’s phrase), the agent not only expresses an objectionable attitude, but also indicates a bad character. In part this character stems from his willingness to embrace and identify with that attitude. The problem is not merely that the agent holds an objectionable attitude, which is in itself blameworthy. Rather, the problem is that the agent, in consciously acting, seems to accept this attitude as his own and reveal something important about himself. This gives the attitude a deeper significance, as the agent embraces the objectionable attitude as indicative of his agency rather than simply expressing it.

Contrast this sort of malicious agent with one who unreflectively and unintentionally steps on my hand because she is too self-involved to pay appropriate attention to her surroundings. This sort of agent seems to reflect in her behavior a lack of regard for those around her, but this lack of regard is different from, and less significant than, a malicious contempt for others. This unreflective hand stomper still reflects something significant about her agency and her character, but she doesn’t express the more positive contempt of the former agent. Her act has significance, and the attitude has significance, but it is simply not as objectionable as the agent who consciously and reflectively stomps on my hand. This more contemptuous agent not only expresses the objectionable attitude, but embraces it.

If this is right, then agents who consciously and reflectively engage in objectionable behavior reflect more objectionable attitudes of greater significance. How blameworthy an agent is for some attitude is proportional to the significance of the objectionable attitude reflected in his behavior. Agents who consciously and reflectively engage in objectionable behavior, then, are more blameworthy for such behavior than agents who unconsciously or unreflectively engage in objectionable behavior. This is not to say, of course, that such unreflective agents are not blameworthy. On the contrary, their behavior reflects an objectionable attitude and they are responsible for this attitude. But the attitude their behavior reflects is less significant, and therefore less morally objectionable, than their counterparts who consciously and reflectively engage in such behavior.

Holly Smith floats a related idea, though she never fully commits to it. She writes: “Probably there is a continuum here: we should speak instead of degrees of blameworthiness,
where an agent is more blameworthy if a substantial portion of his psychology was engaged in a decision, and less blameworthy if a less significant portion of his psychology was engaged” (2011: 138). Of course, for Smith, the threshold remains higher than one attitude, but there seems no good reason to draw the line so high. Given the scalar nature of blameworthiness, it is open to RCV to embrace a lower threshold of responsibility while employing the notion that blameworthiness comes in degrees and correlates with the significance of the attitude reflected in one’s behavior.

7.3.2 One Omission Isn’t Enough

Even if one can be responsible for only one attitude, there is concern that one cannot be responsible for forgetting, failing to notice, and negligence because such cases may not reflect any negative attitude or judgment.

Recall Birthday. Angela failed to fulfill her promise to call Dave on his birthday because she forgot about Dave’s birthday. Smith claims that in such a case, Angela’s failure to remember her promise to Dave or to remember Dave’s birthday suggests that Angela doesn’t care enough about her friend or her promises to her friend. But Levy disagrees:

[T]he connection between omissions and attitudes is far too weak to license the inference from failing to recall Z to insufficient caring about Z. Since lapses are caused by attitudes that need not have any content referring to the omission, in claiming that a lapse is evidence of a lack of concern, we are claiming that agents ought to be such that relevant thoughts ought just to have come to mind unbidden. While it is true that (say) the more I care for a friend, the higher the probability that the thought that today is her birthday will occur to me, attention to how this actually works shows that an inference from lapse to lack of care is unjustified. (2011: 251-252)

While it would be a wonderful state of affairs if agents always had the proper rational connection between what they care about and what thoughts they have, the fact is that unfortunately, we are not ideally rational agents. If an ideal agent were to forget to call her friend on his birthday and thereby break a promise, we could justifiably infer that she did not care sufficiently about her
friend. But in imperfect creatures like us, many other obstacles might get in the way, impacting the rational connection between what we value and our thoughts.

As Levy tells the story, the more one cares about some individual, the more she may think about him. She may see cues in the environment that cause her to think of him as well (252). The more she cares for the person, the greater the variety of cues that may trigger thoughts about him. Suppose that Angela fails to recall Dave or the promise she made to him. From this can we conclude that she fails to care sufficiently about Dave? No. While that may have been the cause, she may simply have gotten distracted. Or she may have failed to encounter the appropriate cues in her environment. Because these are genuine possibilities, we can’t infer that Angela doesn’t care sufficiently about Dave simply because she failed to remember her promise on his birthday.

This point generalizes, of course. RCV, like RRV, claims that agents are responsible for failures to notice, forgettings, and negligent behaviors or omissions in virtue of the attitudes such failures, forgettings, and negligent behaviors or omissions reflect.\(^{117}\) But we cannot infer any attitude from some failure or forgetting.

Levy is correct. But notice that Levy’s concern does not really affect RCV. There are two questions we can ask: (1) Did the failure, forgetting, or negligent behavior or omission reflect some attitude? (2) Are we justified in inferring that the failure, forgetting, or negligent behavior or omission reflected some attitude? Levy’s concern is relevant only to (2), not to (1). Even if we are not justified in inferring that the failure, forgetting, or negligent behavior or omission reflected some attitude, it may still be the case that the failure, forgetting, or negligent behavior or omission did in fact reflect some attitude. And this is all that RCV requires. When some failure, forgetting, or negligent behavior or omission reflects some attitude, such as an objectionable attitude or a lack of care, the agent is responsible (and blameworthy) for that failure, forgetting, etc. Regardless of our epistemic position, some behavior either reflects an attitude or it doesn’t. RCV is concerned with the actual reflection of some attitude, not with our epistemic shortcomings regarding whether some attitude has been reflected in some behavior. In

\(^{117}\) To be more precise, these failures, forgettings, and negligent behaviors and omissions might reflect some attitude, but they might also reflect a lack of some attitude. When Angela fails to keep her promise to Dave, this is meant to be an indication that Angela lacks the appropriate attitude of care toward Dave. For conciseness, I mention only reflections of attitudes in the text, but the reader should feel free to supply “or lack of attitude” in each instance.
other words, even if we are not certain whether we are justified in judging an agent responsible for some attitude, the agent may nevertheless be responsible for that attitude.

7.3.3 An Impractical View

Of course, this response raises a further objection. We are rarely, if ever, in a position to justifiably believe that an agent is responsible for some failure, forgetting, or negligent behavior because we are rarely, if ever, in a position to justifiably believe that the failure, forgetting, or negligent behavior reflects some attitude. In other words, we are left with an impractical theory that is difficult to apply to a significant range of cases. Insofar as a theory of responsibility ought to be practically useful, RCV clearly falls short.

While there is some difficulty in applying RCV to some cases because of our limited epistemic position, this difficulty shouldn’t be overstated. The difficulty in applying the theory seems limited to one-off cases of forgetting, negligence, etc. In cases of intentional wrongdoing, we are justified in inferring that the agent has some objectionable attitude and that the attitude is reflected in that action or omission.

From one instance of forgetting a birthday, we may not be justified in concluding that Angela doesn’t care about Dave (or doesn’t care enough about Dave). But our evidence becomes much stronger when it comes to patterned behaviors. Even Levy admits as much when he writes that “[w]e can justifiably infer from a relatively consistent pattern of lapses, on the part of a normal agent, to his or her attitudes toward the objects that he or she neglects” (2011: 252). If an agent repeatedly and consistently fails to notice some feature, or if she forgets her promises, we can infer that she does not care sufficiently about some person, some feature, or her promises.

The reason for this, according to Levy, is that agents typically get significant reinforcement when they repeatedly forget promises or fail to notice something. This reinforcement can raise an agent’s probability of noticing cues in the environment in the future. But more importantly, “[t]he longer the period of time, and the more conducive the internal and external environment, the lower the probability that my failure to recall is a product of my bad luck rather than of my failure to care sufficiently” (253). In a one-off case of forgetting, the chance that one will encounter environmental cues to remind one may be fairly low depending on the environment. Over time, however, the probability that some cue will occur such that the appropriate thoughts come to the agent’s mind increases. If a pattern of forgettings takes place
over a longer period of time such that a person frequently and regularly fails to keep her promise, we may have sufficient justification to hold her responsible for the failure in virtue of the justified inference that she has or lacks some attitude.

How patterned some behavior must be, or how often some lapse must occur for us to be justified in inferring an objectionable attitude on the part of the agent, is a difficult question without a clear answer. It will simply be difficult to tell in certain cases whether an agent’s behavior or omission reflects some attitude or other. I think this is a virtue of RCV, however. There are cases in which it isn’t clear that an agent is responsible for some item, and we are uncertain in our responsibility ascriptions. Often, these cases are just the sort described, where an agent fails to fulfill a promise or omits to do something one time. Only when we see a pattern of such failings do we begin to step out of our uncertainty and feel justified in holding the agent responsible for such failings.

While we would ideally like to know in every case when agents are responsible or not, the fact that we are epistemically limited is not a difficulty unique to RCV. Even volitionists will, at times, have difficulties in showing that an agent is clearly responsible for some action or omission. It is difficult to know whether an agent has acted or omitted intentionally at certain times, but we certainly do not think that we therefore should abandon volitionism as deeply problematic, useless, or impractical.

The moral is this: while it may be difficult to know whether some behavior or omission reflects some attitude in a limited range of cases, such cases do not mean that we ought to reject RCV. In cases where we are not justified in inferring responsibility, the best practice is to adopt a principle of caution. An agent may deserve blame, according to RCV, and yet we might refrain from blaming the agent due to our lack of justification. Nevertheless, when a failure, forgetting, omission, or the like reflects some attitude, the agent is responsible for that failure, forgetting, omission, etc.—even if we aren’t in a position to know it.

7.3.4 What About Punishment?

Thus far, I have been discussing responsibility that warrants the reactive attitudes. On RCV, agents can be directly responsible for their attitudes, which means that they can be blamed

---

118 Matthew Talbert makes a similar point (2013: 235-237).
for their objectionable attitudes. But responsibility is closely tied with sanction and punishment. As Levy writes,

[A]n agent is morally responsible for an action or an omission if the fact that they have performed that action, in the circumstances and manner in which they acted, is relevant to how they may permissibly be treated when it comes to the distribution of benefits and burdens. To say an agent is blameworthy for an action, for instance, is to say that (ceteris paribus), because they have acted in that way, they may permissibly be punished or that, if burdens are to be distributed, it is better that they fall on them rather than on others who are not blameworthy. (2014: 2-3)

On this understanding of responsibility (or more accurately, blameworthiness), agents who are blameworthy may be punished as a result of performing some wrong action or omission. As this applies to nonvoluntary items like attitudes, such an understanding of responsibility seems more substantial than the nonvolitionist had in mind. To say that an agent is blameworthy for an objectionable attitude, one might think, is not to say that because they have that attitude, they may permissibly be punished. Rather, the focus is simply on the reactive attitudes, as I claimed in the first chapter.

It looks like the advocate of a nonvolitionist view such as RCV is faced with a dilemma: either agents can be punished for their objectionable attitudes or they cannot. But it seems counterintuitive and unfair that agents can be punished for their objectionable attitudes. If, on the other hand, agents cannot be punished for their objectionable attitudes, then it seems that the advocate of RCV is not talking about the same kind of responsibility as the volitionist. Instead, the two are simply talking past each other.

Many nonvolitionists seem to grasp the second horn, as they attempt to drive a wedge between blame and punishment. In addressing the charge that the reactive attitudes are unfair to impose on those with objectionable attitudes, Pamela Hieronymi writes:

---

119 Levy focuses on actions and omissions here because, as a volitionist, he thinks that agents are directly responsible only for actions and omissions.
The idea that resentment can be unfair because its target did not have an adequate opportunity to avoid it gains its power in part by thinking of resentment (and other such responses) as a penalty, sanction, or punishment that we intentionally impose on wrongdoers. … But to think of resenting as imposing a penalty or sanction is, I would argue, to confuse certain nonvoluntary reactions, such as resentment, with intentional, punitive activities like guilt-tripping (trying to make someone feel bad about what he or she has done). … Once we set aside guilt-tripping and other intentional, punitive responses to wrongdoing, and focus, instead, on the nonvoluntary changes in attitudes and relationships that typically accompany the negative actions and attitudes of responsible people, I believe the appeal to fairness loses much of its power. (2014: 29-30)

Hieronymi here is setting aside considerations of punitive sanctions and punishment and focusing purely on the reactive attitudes. She is speaking only to the appropriateness of the latter, not of the former.

Angela Smith makes a similar distinction: “[T]here is a distinction between judging a person to be responsible and culpable for her behavior, on the one hand, and engaging in various forms of blaming activity (including punishment), on the other” (2008: 377). Here, Smith seems to be distinguishing between judging blameworthy and blaming. This is an important distinction, as blaming, unlike judging blameworthy, fundamentally involves a reactive attitude. The distinction at issue here, however, is between blame and punishment. If Smith means to refer to blame as punishment here, then she is making the same point that Hieronymi makes above.

120 Smith says, “judging a person to be responsible and culpable,” but I take it that this must be interpreted as judging blameworthy in this particular instance. While being responsible for A does not mean that one is blameworthy for A, this is because A might not violate any moral norms such that it is wrong or bad. In such a case, one might be responsible for A but not blameworthy for A. I take it that this is not the sort of case Smith has in mind, however. Nevertheless, Smith does seem to hold that there are further conditions on blameworthiness apart from those on responsibility for bad or wrong items (Smith n.d.). I argue against further conditions on blameworthiness apart from those on responsibility for wrongdoing in Fritz 2014.

Alternatively, Smith may be distinguishing between blaming and engaging in blaming behavior. Perhaps having a blaming attitude is a variety of blaming, though one need not express the attitude by engaging in blaming behavior, and so one can blame without engaging in blaming behavior. In this case, Smith may understand engaging in blaming behavior to be punitive, though blaming simply by having some attitude may not be punitive.

121 If Smith does not see blame as punitive, then I take it that the dispute simply shifts to the nature of blame. Many who see punishment as closely tied to responsibility do so in part because they understand blame as punitive. If blame is not punitive, however, then perhaps the nonvolitionist can avoid entering the dispute regarding punishment and responsibility altogether. The conditions on punishment, the nonvolitionist might insist, are altogether different from the conditions on blame, as the former involves harms, while the latter is simply a form of moral protest (see Smith 2013). The differences in conditions have nothing to do with the conditions on responsibility, however, nor do
While he does not identify as a nonvolitionist, Michael McKenna also illustrates this distinction between blame and punishment and offers reasons why the two should not be identified (2012: 134-146). Although blame often harms others by compromising their capacities to engage normally in social intercourse and be emotionally stable, punishment harms in a much wider range of ways. Also, in blaming—either overtly toward the agent or by simply holding an unexpressed blaming attitude—one need not intend to harm another, even if that harm is an inevitable consequence of blaming. In contrast, something cannot be punishment without being intended to harm the individual. So Rita might mildly blame Sid without intending to harm him, causing him only minor social harm. This is a case of blame without punishment. In contrast, Rodney might find his employee Sean incredibly charming and difficult to blame. Nevertheless, as a result of Sean’s behavior, Rodney may fire Sean as a form of punishment without blaming him. While punishment and blame are closely related, each can exist without the other.

The nonvolitionist typically appeals to this distinction to show that blaming attitudes are appropriate and fair, even if the individual could not have avoided the blaming attitudes (see 5.3.2). Hieronymi claims that it is not unfair to distrust someone who is notoriously unreliable, even if the person cannot make herself more reliable or if she lacked the opportunity to do so in the past. Similarly, it is not unfair to resent someone who is impatient or rude, even if the person cannot now (or could not in the past) make herself more patient or kinder. “The attitude…responds, simply, to the disregard shown to one person by another. It is the negative reaction that marks that fact in one offended” (Hieronymi 2014: 32). The negative reactive attitudes are appropriate responses to faults in others, and they do not carry with them the requirement that the person could have done better or could have avoided acting some way or having some attitude (Hieronymi 2004 and 2014: 30-31). Thus, it is not unfair to respond with blame toward those who have objectionable attitudes, even if those attitudes could not have been avoided.

Drawing this distinction, however, is not entirely helpful here. The volitionist might happily accept the distinction between blame and punishment. But it’s important to see that blame can be used as a form of punishment, and perhaps it often is. One might think that blame
for attitudes is justified, but deny that blame as punishment for attitudes is justified or fair. After all, the blame is intended to harm the individual with those attitudes, simply in virtue of the agent’s having those attitudes.\textsuperscript{122} It seems incredible, however, that one could be justified in harming someone simply for having objectionable attitudes. Such a line insists that it is, in principle, inappropriate or unjustified to punish agents for having objectionable attitudes. Can the nonvolitionist make this claim?

Nonvolitionists often simply avoid this question altogether, focusing instead as Hieronymi does on whether the reactive attitudes are justified rather than punishment. (I myself sidestepped the question in chapter 1.) But there is a deep worry here: If nonvolitionists insist that agents cannot, in principle, be punished for their attitudes, then this suggests that punishment itself has stronger requirements than the nonvolitionist can accommodate. It seems that justified punishment presupposes responsibility of a certain sort, and if punishment is not, in principle, justified for attitudes, then this may suggest that the nonvolitionist is not interested in a kind of responsibility that is robust enough. In such a case, nonvolitionists and volitionists may, in fact, be talking past each other.

If nonvolitionists insist that justified punishment requires more control or a different sort of control than rational control, then I agree that nonvolitionists and volitionists do seem to be talking past each other. The nonvolitionist will be impaled squarely on the second horn of the dilemma above. The volitionist will insist that a kind of responsibility that does not require voluntary control is toothless, as it cannot justify punishment. The volitionist is concerned with responsibility that justifies blame, punishment, and blame as punishment. Whatever kind of responsibility the nonvolitionist is concerned with, one might protest, it isn’t the same kind of responsibility as that with which the volitionist is concerned.\textsuperscript{123}

\textsuperscript{122} Here I assume a desert-based, retributivist understanding of punishment rather than a consequentialist one. If whether we should punish is a matter of consequences, I take the response of the nonvolitionist to be rather straightforward: we should punish an agent for bad attitudes if and only if punishing the agent would have better consequences than not punishing the agent. Perhaps punishing the agent would cause the agent not to have those attitudes, or to revise those attitudes, or to take steps to remove or mitigate those attitudes. If that’s the purpose of punishment, then it seems clear that such punishment is justified and appropriate. The concern at issue, however, is one of desert and fairness, which the consequentialist picture does not address.

\textsuperscript{123} There is a response open to the nonvolitionist here. Perhaps there are multiple kinds of responsibility at issue, such that volitionists are concerned with a more robust sense of responsibility than nonvolitionists. (As I discussed in chapter 1, perhaps volitionists are concerned with responsibility as accountability, and nonvolitionists are concerned with responsibility as answerability.) This is a concession to the volitionist, to be sure, and most nonvolitionists will resist it, I believe (see Smith 2012 and Talbert 2012a). But the nonvolitionist might claim that, given the arguments of the proceeding chapters, there is good reason to think that the volitionist is not justified in
While nonvolitionists rarely address the problem of punishment for attitudes, one might hold that punishment requires further conditions, but not further control conditions. Angela Smith suggests something like this response. She argues that if we are focused on accountability as requiring the justification of further punishment responses such as “official public censure and other forms of punishment,” then we are seeing accountability blame as “akin to legal punishment” (2008: 378). It will involve “active blaming responses of various sorts, including both public censure and punishment as well as active expression of negative attitudes such as blame, criticism, resentment, and indignation” (379). But, she continues,

[I]t is important to recognize that “accountability blame,” interpreted in this way, is a type of activity that is governed by a variety of moral and non-moral norms, a great many of which have nothing to do with the basic conditions of moral responsibility. Our judgments about when it would or would not be appropriate to actively express moral criticism or to punish someone for some action or attitude are sensitive to a host of considerations that have little or nothing to do with that person’s responsibility and culpability for the action or attitude in question. These considerations can include, among other things, our relation to the person, our stake in the matter, the significance of the fault, and the person’s own response to her failure. We may conclude on the basis of any of these things that blaming activities and responses would be inappropriate in a given case, without thinking that the person is therefore not responsible (or not “open” to moral criticism) for the action or attitude in question. (379)

Smith’s thought is that formal sanctions are not justified only when an agent is responsible for some item. Instead, their justification depends on a whole host of other considerations that are independent of the conditions of direct responsibility. So simply because punishment is not justified for attitudes, we are not warranted in inferring that the agent is not responsible, or that responsibility requires further conditions, such as greater control or control of a different type.
I think Smith’s move here is an important first step for nonvolitionists. But notice that it does not go far enough, because it still does not address the question of whether punishment for attitudes is, in principle, inappropriate. Suppose that Smith’s further conditions above are satisfied in the case of an individual with objectionable attitudes: we are related to the person in the right way, we have an adequate stake in the matter, the attitudes are significant, and the person does not respond appropriately to the attitudes. (We might add in further conditions here, as I discuss below.) Could the person thereby, in principle, be punished for these attitudes? For all Smith has said here, the answer could be affirmative.

In fact, I think the right answer for the nonvolitionist here is an affirmative one. This is to grasp the first horn. But the nonvolitionist shouldn’t stop here; instead she should attempt to dull the sharpness as much as possible. In what remains of this section I’ll attempt to do just that.

The right way to dull the sharpness of the first horn, I’ll argue, is to allow that punishment for attitudes is, in principle, justified, but to also claim that punishment that goes beyond blame is rarely, if ever, in fact justified. The reason for this is that these further conditions on significantly harmful punishment will rarely be satisfied in the case of objectionable attitudes. Smith lists some of these conditions above, but I’d like to highlight two here: (1) the degree to which the offense is bad or wrong, and (2) the epistemic justification we have in thinking that the person is guilty.

First, how much punishment one deserves for some offense depends on the nature of the offense for which one is guilty. Punishments must be appropriately proportional to offenses. It’s plausible that most objectionable attitudes aren’t bad enough to warrant much, if any, punishment. Simply having an objectionable attitude might cause some harm to the individual, but for the most part its negative consequences are minimal. Just as a minor wrongdoing warrants only minor punishment, most objectionable attitudes are minor offenses. It seems that the most punishment they could warrant would be blame as punishment. But as I have argued, the harms associated with blame as punishment are minor. Additionally, the attitudes for which the agent is blamed are under the agent’s rational control. The agent expresses her values in those attitudes. When we punish agents for actions or omissions, we punish them at least in part in virtue of the values those actions or omissions reflect. We care about these actions or

---

This is not to say that having objectionable attitudes does not manifest in wrong behavior or result in negative consequences. On the contrary, as I discuss below, this will often be the case. Instead, the claim here is that merely having some objectionable attitude has few, if any, bad consequences.
omissions because of their bad qualities, but also because they show us something bad about the agent performing them. We frequently cite an agent’s quality of will as being relevant to how we ought to treat her. And an agent’s attitudes are partially representative of the agent’s quality of will. So it is not completely implausible that punishment for attitudes could be justified, provided that one keeps in mind that only minor punishment, such as blame as punishment, could be justified.

The second condition, however, is a powerful reason why punishing agents for merely having attitudes is rarely, if ever, practically justified. We simply do not have access to the attitudes of other agents. In order to be justified in punishing an agent for some item, however, we must have strong evidence that the agent is actually guilty for that item. The reason we need this justification is due to the harms and burdens that come with punishment. Because these harms are more substantial than the harms associated with mere blame, we need much more justification that an agent is guilty before we are justified in administering those harms in the form of punishment. Because we lack insight into others’ attitudes, however, we simply lack this epistemic justification.

This insight can help to explain why we might be inclined to think that agents must have some kind of voluntary control over something in order to be punished for it. Voluntary control is strong evidence that an agent has some values or makes some judgments. The agent might choose to act in a particular way, and this choice and this action will reflect the agent’s values and judgments. If an agent doesn’t have this control over her behavior, we think she is excused from responsibility for the behavior and consequently cannot be punished. But there is a simple nonvolitionist explanation for this: in some cases, agents who lack this sort of control do not express any values or evaluative judgments in their behavior. If I fail to pick you up from the airport because someone has paralyzed my body, my omission does not reflect some objectionable attitude toward you or some lack of care about you. So voluntary control can sometimes be useful evidence for determining whether an agent has an objectionable attitude. When an agent intentionally behaves wrongly, there is strong evidence that the agent has an objectionable attitude.

On the other hand, rational control need not be visible to anyone at all. An agent might have a multitude of objectionable attitudes, but we might not have evidence for these attitudes. Of course, sometimes we will get strong evidence in the form of intentional actions and
omissions. And sometimes we will get some evidence in the form of nonvoluntary behavior, such as an unintentional omission, a negligent omission, or a nonvoluntary scowl or smirk. But such evidence is fallible—too fallible to stand as evidence for doling out the significant kinds of harms associated with punishment. As we have seen, just because an agent forgets her dog in the car, for instance, we cannot infer that she has an objectionable attitude toward her dog or that she doesn’t care sufficiently for her dog.

This means that voluntary control has important connections to punishment. It is useful in providing the evidence that justifies punishment. While voluntary control isn’t required for punishment, it is helpful in providing epistemic assurance that the agent has some objectionable attitudes. But we only have rational control over our attitudes, not voluntary control. So one of the best methods we have for determining whether an agent has objectionable attitudes is unavailable to us, and our epistemic certainty is significantly undermined.\footnote{In personal discussion, Daniel Miller suggested that the epistemic restraint might be satisfied here if the agent were to simply report his attitudes. Suppose Dan tells me that he has an immense amount of hatred for me. Am I justified in punishing him for this hatred? We might have more justification, but it’s not clear that this is enough. Agents are not always honest, and they do not always know themselves as well as they think they do. Dan might be mistaken about his own attitudes, for instance. But even if this does provide enough justification, the first condition above will come into play. Dan’s hatred towards me is bad, but it’s not clear that it is bad enough to warrant anything more than blame (including blame as punishment).}

To summarize, punishment for attitudes is justified in principle because such attitudes are under one’s rational control. This is to grasp the first horn of the dilemma above. But any stronger punishment than blame for attitudes is not practically justified because these attitudes are not significant enough faults and we lack the epistemic justification needed to intentionally impose significant harms on an agent.

The discussion of epistemic justification and badness might prompt a further question, however. Sometimes, objectionable attitudes manifest in nonvoluntary behavior that causes significant harm and suffering. And it also seems that we have more epistemic justification in attributing some objectionable attitude to an agent if she engages in some bad behavior, even if that behavior is nonvoluntary. So even if agents frequently can’t be punished for simply having attitudes, it seems they might be justifiably punished, in practice, for nonvoluntary behavior.

I think this conclusion is warranted. We are more justified in punishing agents for nonvoluntary behavior than we are for merely having certain attitudes. But note several things about this claim. First, we have moved from a discussion of punishment for \textit{attitudes} to punishment for \textit{behavior}. When we punish for nonvoluntary behavior, we are punishing the
behavior, not punishing merely for having some attitude. Second, we do, in practice, sometimes punish people for nonvoluntary behavior. Agents are held to account for their negligence, for instance. Third, agents will not be justifiably punished for all nonvoluntary behavior, but only behavior that reflects an objectionable attitude, is significant enough to warrant the intentional imposition of significant harms, and for which we are justified in inferring an objectionable attitude. Additionally, there are further conditions that apply, including those that Smith lists above. So even if it is justifiable to punish an agent for nonvoluntary behavior, it is still an open question of how much punishment is appropriate and whether any individual instance of nonvoluntary behavior warrants punishment. Lastly, the claim that agents can be justifiably punished for nonvoluntary behavior is consistent with the claim that agents who intentionally or voluntarily behave wrongly or badly should be punished more severely. The reason for this is that an agent’s quality of will is relevant to the sort of punishment she deserves, and RCV takes this sort of consideration into account.

If the above arguments are persuasive, why have nonvolitionists been shy to grasp the first horn openly? One reason might be because of the concern that such a position is highly counterintuitive, and nonvolitionists often claim that they are offering a position that is intuitive and matches our everyday responsibility ascriptions. But when we take into account that we are rarely justified in actually punishing agents for having attitudes, and that the punishment warranted is often only as strong as blame, the counterintuitive worries dissipate. The first horn of the dilemma is not nearly as sharp as it first seems.

The general point is this: Agents can, in principle, be punished for their objectionable attitudes and nonvoluntary behaviors that are wrong or bad. But in the case of attitudes, the practical requirements for justified punishment will rarely, if ever, be met. When they are, the punishment warranted will be only blame. The charge that RCV and other nonvolitionist views are not talking about a robust type of responsibility because they do not warrant punishment can be adequately addressed.

7.4 Applying RCV

With a clearer view of RCV, let’s apply the view to the cases we first explored in chapter 1. Here are the cases once more:
**Sore Tooth:** Fiona, a politically liberal person who claims that the color of a person’s skin does not affect a person’s worth, has a toothache and seeks a recommendation for a dentist from a friend. Upon entering the dentist’s office and meeting her new dentist, Fiona feels a sudden pang of dismay when she realizes the new dentist is an African-American. (adapted from H. Smith 2011: 140)

**Fellowship:** Dan and John have both applied to the same fellowship. While both applicants have worked hard in their careers, John has put in more work than Dan has and is more deserving of the fellowship. When it is announced that John has won the fellowship, Dan feels resentment well up inside of him, which manifests in a scowl.

**Birthday:** Angela promised her close friend Dave that she would call him on his birthday. Despite her promise, Dave’s birthday comes and goes, and Angela forgets to call. (adapted from A. Smith 2005: 236)

**Bad Joke:** Ryland is very self-absorbed. Though not malicious, she is oblivious to the impact that her behavior will have on others. Consequently, she is bewildered and a bit hurt when her rambling anecdote about a childless couple, a handicapped person, and a financial failure is not well received by an audience that includes a childless couple, a handicapped person, and a financial failure. (from Sher 2009: 28)

When I first presented these cases, I claimed that they were underdescribed. Now we are in a position to see precisely how.

Consider **Sore Tooth** first. In order to know whether Fiona is directly responsible for her pang of dismay, we need to know whether there is some judgment that the pang of dismay reflects, and whether that judgment was formed via her rational processes or whether its formation bypassed those rational processes. While it is open that Fiona judges that African-Americans are inferior in some way that affects their ability to provide dental care, for instance, it is plausible that Fiona’s pang of dismay does not reflect any morally objectionable attitude or judgment. Instead, it likely reflects merely some associations that are not under her rational control. We might even suppose that she is aware of her implicit biases and has worked hard to
attempt to eradicate them. Despite her best efforts, she may have failed to eliminate or bypass such biases. If Fiona has done all we could reasonably expect her to do in an attempt to eliminate her implicit biases, yet she still harbors some of these biases, she is not blameworthy, according to RCV. If, however, she became aware of these biases and yet took no steps to correct for them, she is responsible and blameworthy for these biases—though only derivatively. She cannot be directly responsible for these implicit biases because they do not satisfy RCd or HC. They are not under her rational control, were formed in ways that bypassed her rational processes, and do not reflect any objectionable attitude.

Next, consider Fellowship. Again, the details of the case matter, and the details are what are often missing in such cases (both in philosophers’ examples and in cases we encounter in the real world!). Nevertheless, it is plausible that Dan reflects some objectionable evaluative judgment in his scowl. The scowl seems to reflect Dan’s resentment of John and his disposition to regard John in negative ways. But the scowl does not directly reflect Dan’s evaluative judgments or values; only his attitudes can do this. Dan’s resentment reflects his evaluative judgments and values, and Dan’s scowl indirectly reflects those judgments and values in virtue of reflecting his resentment. Dan’s resentment is sensitive to his reasons. If he judged that John deserved recognition for good work and that John’s work was superior to his in this instance, perhaps he would not have this resentment. So Dan is directly responsible and blameworthy for his resentment, but he is only derivatively responsible and blameworthy for his scowl—though this derivation involves only a very short trace. Of course, scowling is a minor offense, so Dan’s blameworthiness here is likely minimal.

Whether Angela is responsible in Birthday depends on further details of the case. Does Angela’s failure to keep her promise to Dave reflect an objectionable attitude? Does it reflect that she doesn’t care sufficiently about Dave or her promises to him? If so, then she is responsible and blameworthy for failing to keep the promise. But it’s important to keep in mind that Angela’s failure to keep her promise might not reflect any objectionable attitude. Perhaps she simply forgot. As I argued in 7.3.2, we cannot judge simply from the fact that Angela failed to keep her promise that her failure reflects some objectionable attitude. If this was a one-off case, Dave ought to be understanding and cautious in his blame. If Angela frequently forgets her promises to Dave, however, Dave has more evidence to think that Angela simply does not care
about him, and Dave is more justified in holding her blameworthy for her failure to fulfill her promises to him.

Lastly, Ryland does not seem to be directly responsible for telling an offensive joke or for offending people in *Bad Joke*, as direct responsibility is limited to attitudes. Nevertheless, she can be directly responsible for the attitude or attitudes that caused her to tell the joke and that are reflected in the joke. Her anecdote reflects an attitude of self-absorption and inconsiderateness. This lack of care of others is itself objectionable, and the attitude that reflects it is therefore directly blameworthy. Although Ryland did intentionally tell a joke, she did not intentionally tell an offensive joke or intentionally offend people. Nevertheless, RCV maintains that Ryland is derivatively responsible for telling the offensive joke and for offending people because these reflect her objectionable attitude, and she is directly responsible for this attitude.

There is much work to be done yet in clarifying RCV and its commitments. Still, I hope to have motivated the view and to have shown that we have good reason to take it seriously. RCV can explain how and why agents are responsible for intentional actions and intentional omissions, as well as attitudes and some unwitting omissions or instances of negligence. It can do so in a unified way that seems to match many or most of our ordinary responsibility ascriptions and practices. The fact that RCV can better account for our responsibility ascriptions and practices than VT2 or VT4 is a strong reason to favor RCV over those volitionist views that Angela Smith calls prior choice views. The fact that RCV offers a grounding in the agent’s evaluative judgments and so builds in an appeal to the agent’s quality of will is a strong reason to favor RCV over the capacity views captured in VT3. But RCV not only stands as a strong alternative to volitionist views; it is, I have argued, the most plausible nonvolitionist view available. There is good reason to take the view seriously and develop it further.
REFERENCES


BIOGRAPHICAL SKETCH

Kyle Fritz grew up in Galesburg, Illinois. He received a Ph.D. in Philosophy from Florida State University, an M.A. in Philosophy from the University of Florida, and a B.A. in Philosophy and History from Millikin University. His primary research interests concern issues related to the nature and justification of moral responsibility as well as the nature and norms of blame—including the ethics of blame.