ON THE HOOK: RESPONSIBILITY IN REAL LIFE

Samuel Reis-Dennis

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Philosophy.

Chapel Hill
2017

Approved by:
Susan Wolf
Thomas E. Hill, Jr.
Douglas MacLean
Geoffrey Sayre-McCord
Rebecca Walker
ABSTRACT

Samuel Reis-Dennis: On the Hook: Responsibility in Real Life
(Under the direction of Susan Wolf)

Under what conditions are we responsible for what we do?

Control theorists (e.g. Fischer, McKenna, and Pereboom) hold that we are responsible only when what we do is under our control. Self-disclosure theorists (e.g. Scanlon, Strawson, and Arpaly) hold that we are responsible only when what we do reveals something significant about us, our judgments, our cares, or the quality of our wills. I reject both views in defending an “ecological,” approach to responsibility that I develop by examining cases in which angry blame is an apt response to agents’ behavior.

I argue that expressions of resentment are fitting when they are sanctioned by norms of a justified blaming practice. I argue that our blaming practices are leveling mechanisms that respond to socially disruptive wrongdoing. Since such wrongdoing requires neither control nor self-disclosure, control and self-disclosure theories are both false. In fact, I argue that all reductive accounts of responsibility distort our conceptions of blameworthiness and blameworthy agents.

I propose to understand the conditions of responsibility in general, and blameworthiness in particular, as irreducibly social. This method, with its shift of focus, helps explain and justify the expression of anger and resentment in the cases that serve as counterexamples to control and self-disclosure views. Uncontrolled behavior can be disrespectful, and agents can non-disclosively fail in their obligations to others. In these cases, victims’ expressions of resentment can foster self-respect and prompt apology, without which they may rightly feel diminished.
For Nate, Jeremy, Gruber, Yarbs, Tyler, Jacob, and Ben, the self-proclaimed Most Blamed Man in America.
ACKNOWLEDGMENTS

I had never heard of the philosophical idea of moral responsibility until, as junior at Cornell University, Erin Taylor allowed me, an undergraduate she did not know, to take her graduate ethics seminar. A paper I wrote for that course, on T.M. Scanlon’s account of blame in *Moral Dimensions*, became the basis of an honors thesis I completed the following year under the supervision of Derk Pereboom, who, through his generosity and encouragement, nourished my love for philosophy and inspired me to attend graduate school.

After those great experiences, arriving at UNC only to realize how little I knew was intensely humbling. I am overwhelmed reflecting on how much I’ve learned from my teachers here, and on the depth of their positive influence on my philosophical thinking. I am so grateful to my committee members for their wisdom, diligence, and compassion. I owe the greatest debt to my advisor, Susan Wolf. Knowing that I wanted to work with Susan, and fearful of making a fool of myself, I steered completely clear of her until my third year of graduate school. I’m glad I finally mustered the courage to introduce myself, because working with her on this project has been one of the greatest joys of my life. I cannot imagine what kind of philosopher I would be without her example and direction. I am proud to call her my mentor.

I have also been lucky to learn from an outstanding peer group. I have benefitted from, and enjoyed, discussions of this work with Rob Smithson, John Lawless, and others. And I am especially grateful for the friendship and guidance of Vida Yao, who took me under her wing, and, in moments of doubt, renewed my sense of the value of ethics as an academic discipline.
Most importantly, I am appreciative of my friends and family, whom I cannot thank enough for their support and love. Since college, I’ve sent every paper I’ve ever written to my parents and my grandmother; they’ve read and commented on them all. My grandmother, Pamela Reis, deserves special credit. I tend to send her my work in the middle of the night, and “Grams,” an unusually early riser, meticulously edits it the moment she wakes up, usually getting my papers back to me before I’ve even gotten out of bed in the morning. I am lucky to have the benefit of her eagle eye and sense of literary style. Finally, thank you to my wife, Susan, whose steady confidence, brightness, and playful sense of humor enliven every day.
TABLE OF CONTENTS

1 How Not to Argue About Blame................................................................. 1

2 Anger: Scary Good......................................................................................... 17
   I: INTRODUCTION..................................................................................... 17
   II: WHAT MAKES US ANGRY? ............................................................... 19
   III: WHAT DOES ANGER SEEK? .............................................................. 22
   IV: SCARY, GOOD ................................................................................. 26
   V: ANGER AMONG INTIMATES: ............................................................... 27
   VI: ANGER TOWARD STRANGERS ......................................................... 32
   VII: CONCLUSION: TWO PLEAS FOR COMPOSURE ......................... 36

3 Control Freaks.............................................................................................. 40
   I: INTRODUCTION..................................................................................... 40
   II: THE CONCEPTUAL INTERPRETATION .................................................. 41
   III: THE ETHICAL INTERPRETATION ....................................................... 48
   IV: CONTROL, PRACTICES, AND INCOMPATIBILISM .............................. 56

4 Responsibility and the Shallow Self ........................................................... 60
   I: INTRODUCTION..................................................................................... 60
   II: THE SELF IN EVERYDAY LIFE ........................................................... 61
   III: FRANKFURT AND WATSON .............................................................. 65
   IV: RESPONSIBILITY AND THE SHALLOW SELF ..................................... 68

5 How To Argue About Blame........................................................................ 83
I: INTRODUCTION.................................................................................................................................83
II: TWO BARRIERS TO BLAMEWORTHINESS ..................................................................................84
III: BLAME AS A LEVELING RESPONSE TO WRONGDOING .......................................................87
IV: RULES, EXPECTATIONS, EXCUSES ..........................................................................................89
V: CONCLUSION: WIDE AGENCY AND THE SHALLOW SELF ......................................................95
REFERENCES .........................................................................................................................................97
Many disputes in ethics seem to matter, at least in part, because of their relevance to blame. We tell our students that the answers to what might at first glance seem to be obscure, purely philosophical, questions are relevant to their lives by noting their connection to the justification of our blaming practices. The obvious everyday importance of blame has helped to motivate inquiry into the compatibility of free will and determinism, the existence of moral luck, the nature of intentional action, the possibility of genuine moral dilemmas, and the categorical force of moral considerations, among other areas. As such, it seems reasonable for philosophers to take a further step and investigate the nature of blame. How can we determine how, say, moral luck should affect blame without first understanding what blame is? The impulse is understandable, but I will argue that this project is at least slightly misguided.

In fact, I contend that there is no unified essence of blame. Moreover, even if one could somehow isolate such an essence, it is not obvious that discovering it would help us to clarify or answer the moral questions we care about. In this paper, I canvas some recent attempts to analyze blame and explain why the results of these investigations are unsatisfying. Because English speakers use the word “blame” to refer to so many different responses, the search for the boundaries of the concept only serves to narrow the scope of the moral debate about our reactions to wrongdoing. If, for example, the philosopher determines that mere moral disapproval of an agent’s behavior is not officially blame, then consideration of the legitimacy of moral disapproval is off the agenda for that discussion; if she decides that blaming involves a judgment but not necessarily the feeling or expression of negative reactive attitudes, then angry and resentful responses are deemed irrelevant, and pushed to the background. But what one means when he asks whether or not blame is justified can vary considerably. Depending on the context in which the
question is asked, he might have anything from punishment to disapproval of an agent’s instrumental reasoning in mind. In fact, the word “blame” often functions as a catchall meant to provoke consideration of a wide range of reactions to bad behavior. As handy as “blame” can be as a blanket term, its use in moral philosophy has, at times, come at the cost of precision.

I argue that the best way to begin an investigation of the legitimacy of a moral reaction is to specify a case and a response. Whether that response officially counts as “blame” is usually of minimal interest, and, in most cases, will be obvious to competent speakers. In offering this suggestion, I hope to steer the blame debate in a more fruitful direction, one that will allow for genuine progress toward understanding the questions that have prompted philosophers to attempt to define blame in the first place.

One central source of disagreement in the blame literature is the question of whether blame, at its core, is a judgment, or whether the essence of blame is to be found in the feeling and/or expression of a reactive emotion. Disputants on both sides have marshaled ordinary language intuitions to help make their cases. Observing the lack of progress in this debate, another set of recent theorists have offered hybrid accounts of blame that aim to capture the concept’s cognitive and conative elements.¹

I do not wish to survey and criticize every account, but because this paper is dedicated to the correction of a common mistake, I must focus on the work of a few prominent blame theorists. And so I begin with the recent debate between T.M. Scanlon and R. Jay Wallace, two giants of the blame literature: In Moral Dimensions, Scanlon offers an account of blame that he concedes will “no doubt strike some as revisionist.”² Scanlon contends that

To claim that a person is blameworthy for an action is to claim that the action shows something about the agent’s attitudes toward others that impairs the relations that others can have with him or her. To blame a person is to judge him or her to be blameworthy and to take your relationship with him or her to be modified in a way that this judgment of impaired relations holds to be


appropriate.\(^3\)

To illuminate his account, Scanlon imagines learning that, in order to win a few laughs at a party, his friend Joe has betrayed his trust by revealing embarrassing facts shared in confidence.\(^4\) In deciding how to respond, Scanlon outlines three options:

First, I might consider whether I should continue to regard Joe as a friend. An answer to this question is a judgment about the meaning of Joe’s action—about what it shows about his attitude toward me, considered in relation to the requirements of friendship, and about the significance of that attitude for our relationship. Second, I might revise my attitude toward Joe in the way that this judgment holds to be appropriate […]. Third, I might complain to Joe about his conduct, demand an explanation or justification, or indicate in some other way that I no longer see him as a friend.\(^5\)

Scanlon ties these sorts of responses to Joe’s betrayal to his account of blame. The first response (judging that the meaning of Joe’s action warrants a change in the relationship) is a judgment of blameworthiness, while blame itself is tied to the second and third responses (actually revising the relationship on these grounds).

According to Scanlon, this understanding of blame “[f]its the facts of our moral experience […] and the significance it has for us. It also explains various facts about […] the ethics of blame: about who can be blamed, who has standing to blame, and why we should blame.”\(^6\)

For R. Jay Wallace, however, Scanlon’s theory of blame does not capture the facts adequately. In fact, it “leaves the blame out of blame.”\(^7\) Wallace, like Scanlon, holds that blame occupies a kind of intermediate space between cognitive recognition of wrongdoing and punishment. He writes: “It seems clear that to blame someone is not merely to register the fact that they have done something wrong, or

---

\(^3\) Ibid, 128. It should be noted that Scanlon holds that an action that serves to confirm pre-existing negative judgments about an agent’s character is also blameworthy. And along the same lines, blame can consist in taking an agent’s action to confirm (rather than give rise to new) negative character assessments (131).

\(^4\) Ibid, 129.

\(^5\) Ibid, 130.

\(^6\) Ibid, 123.

displayed a morally objectionable attitude. One can acknowledge that a person has violated moral standards in these ways without blaming them for it.” On the other hand, to associate blame too strongly with punishment, Wallace writes, would be to “deprive blame of its legitimacy as a form of interpersonal behavior.”

Wallace diverges from Scanlon in his account of where blame is situated on the continuum between disapproval and punishment. Wallace defends a broadly Strawsonian account of blame, holding that one is blameworthy if he or she is a justified object of negative reactive attitudes like resentment and indignation, and that blaming involves actually feeling the reactive emotions: “My contention is that this expressive connection to the reactive sentiments is the key to understanding the special quality of blame. To count as blaming a person, you have to be exercised by what they have done, and to be exercised in the relevant way is to be subject to one of the reactive sentiments….”

As far as I am concerned, neither a Scanlonian account nor a Strawsonian reactive attitude account of blame actually picks out the “essence” of the concept, or isolates its “special quality.” Much has been written about Scanlon’s theory, and responding to it is not my primary concern in this paper. I will say, though, that I think Wallace’s basic criticism of the account is fair: Scanlon’s conception of blame is too distant from many important uses of the term to plausibly capture its nature or essence. Consider, for example, a “zen” pacifist who responds to a vicious personal attack along these lines: “I’ll never trust or confide in my attacker again, but I never blame anyone.” Here, we have what seems to be a clear case of Scanlonian blame. Is our pacifist making an error? I suspect we all know what he means: roughly, he rejects (at least the negative) reactive attitudes, or at least is not animated by them in this case. This use of “blame” strikes me as quite common. When people wonder whether we ever ought to blame,

---

8 Ibid, 348.
9 Ibid, 348.
10 Ibid, 358. Both Scanlon’s and Wallace’s accounts deserve longer and more detailed treatment. I have no room to offer that here, but I hope that these basic sketches suffice for the general observations I wish to make. The nuances of these two theories, though interesting, are not critical to my main contentions.
they often mean to be calling into question our practice of expressing and feeling negative reactive sentiments.

In fact, consideration of this recognizable examination of blame highlights a puzzling feature of Scanlon’s account: it makes blame too easy to justify. When a dinner table conversation gets “philosophical” and someone provocatively says, “should we ever blame people, though? I’ve found that it always seems to make things worse,” answering that we are obviously justified in, say, ceasing to trust people who have lied to us or in distancing ourselves from friends who spread vicious rumors about us behind our backs, would be unhelpful. I will return to the point that Scanlonian blame is too easily justified later in the paper, but for now I will conclude my brief discussion of Scanlon by claiming he has, at most, highlighted one form of blame and blameworthiness, and probably not the one that immediately springs to mind when most of us hear and use these words.

A reactive attitude account of blame like the one Wallace favors, however, is equally problematic, and for a similar reason. Wallace fails to appreciate that our language sanctions many familiar uses of “blame” that do not imply the feeling or expression of negative reactive attitudes. Imagine, for example, that I had plans to go to a basketball game with a friend and that he stood me up. Suppose I feel no reactive sentiments toward him, but simply explain that I’m sad and disappointed that he bailed, and that I won’t be offering him my extra ticket again any time soon. If he responded by saying, “You can’t blame me for that! My boss made me work late at the last minute!” would he be, strictly speaking, making some sort of mistake? Whether I feel reactive attitudes or not in this case is irrelevant to whether my friend’s response is appropriate and understandable. There is no good reason to deny that the concept of blame is at home in such a response, and thus also in my initial imagined reaction to being stood up for the game.

To see just how fluid our use of “blame” can be, imagine a group of old friends at a high school reunion: One friend confesses that he has been dreading the prospect of running into one particular classmate. When the others ask why, he says, without animus, “I guess I still blame him for what happened between Jacob and me. Thinking about it just bums me out, and I still don’t totally trust him.”
To my ear, this seems like a perfectly reasonable thing for someone who feels no reactive sentiment to say (and think). Still, it would be just as intelligible for the man to say: “Thinking about what happened between us gets me down. I’ve finally let the blame go after all these years, but I still don’t trust him.” I am suggesting that the speaker could feel the exact same way about the classmate in both cases, and that in both cases his speech is likely to be understood perfectly in context. He makes no mistake in either instance. What are we to make, then, of Wallace’s talk of the “nature” or “special character” of blame? What reason, other than a desire for philosophical unity for its own sake, would motivate us to officially rule against either of these everyday uses of the term?

* *

Until now, I have not explicitly questioned Scanlon and Wallace’s shared assumption that blame is neither purely cognitive nor purely punitive. I will argue now that this premise is false. Perhaps the best example of its falsehood involves one of the most interesting and important uses of the word “blame” in our culture: “blaming the victim.” Here is a sketch of a standard case of blaming the victim: A man rapes a young woman at a party. Some people, hearing of the incident, react by thinking and saying things such as “Well, that skirt was really short….” or “She went upstairs to take shots of tequila with him? What did she think was going to happen?” These responses are clear cases of victim blaming. What can we say about them? To be sure, in some cases the people making these statements resent the woman (perhaps they resent women generally) or think that her actions impair her relationships with others. But this need not be, and often is not, the case.

Can’t we rather easily imagine someone who feels no resentment or anger toward the woman at all but says, “she shouldn’t have put herself in that situation by going upstairs without bringing a friend” just the same? This person’s response, though not involving reactive sentiments or the thought that the woman’s attitudes justify others in relating to her differently, is still victim blaming (and is still offensive). To articulate this sort of blame is only to express a judgment; it is offensive not because of any attendant attitude, but rather because it pins “the blame,” stemming perhaps from a judgment of wrongness or faulty reasoning, on the woman, rather than focusing on the actions of her attacker or on the
system that gives rise to a state of affairs in which one must take a friend along at all times in order to be safe at a classmate’s party. Victim blaming, even when divorced from all negative attitudes and judgments about the woman’s character, entrenches the idea that women should dress and behave in certain ways if they wish to be safe. It implies that a man’s responsibility and blameworthiness for committing rape is somehow mitigated if the woman was “asking for it.”

Shifting attention to another cluster of reactions reveals that blame can function as a kind of punishment as well. Consider an instance in which one member of a group of friends acts disrespectfully toward another. Now imagine that one of the friends expresses her indignation by raising her voice and yelling: “Hey! You can’t talk to her like that! Who do you think you are!?” The angry reaction, it seems to me, might aim to encourage apology, reflection, and shame, and this suggests that expression of the reactive attitudes can, at least sometimes, function as a kind of interpersonal punishment: we subject the wrongdoer to something unpleasant in order to get his attention and prompt a certain response. We might also desire to even the score: our friends cannot simply be allowed to act badly and get away with it. Permitting one member of a group to transgress in this way makes the well-behaved members look foolish, even cowardly. They might reasonably desire that their friend feel the pain of guilt for threatening and possibly destroying something of value. Punitive blame invites the wrongdoer to humble himself to the others by expressing contrition and apologizing. By expressing their anger, the friends can accomplish this punitive task, stand up for themselves and their values, and even show the offender that they still see the friendship as salvageable.\(^{11}\)

Or consider a more extreme case from the second season of Bravo’s hit reality series Vanderpump Rules,\(^{12}\) in which a character named Kristen sleeps with Jax, her best friend’s ex. At the reunion show, the host asks Kristen to explain and justify her actions. Rather than attempting a defense, she breaks down in tears and apologizes to everyone. But it’s not enough. The cast members don’t let up.

\(^{11}\) It is important to note that angry blame is not always best understood as a kind of interpersonal punishment. I will address the structure of angry blaming interactions in depth in chapter two.

They yell at Kristen and tell her how disgusted they are by her actions until she is reduced, as the host puts it, to “a puddle of tears.” This case indicates that angry blame need not even aim to give rise to contrition or guilt—it can be so wholly punitive that its function in a given case is largely vindictive, rather than constructive.

In response to this example, Wallace and Scanlon might argue that this sort of blame is never morally justified, that it is too cruel or harsh. This might be true, but I doubt things are so simple. Even in the Vanderpump case, blame allows the friends to save face, clarify their values, and set a precedent for dealing with future wrongdoing. But answering the question whether they are ultimately justified in reacting this way is not my concern here. Rather, I am only suggesting that these reactions intuitively belong in the category of blame and that there are legitimate moral questions to be asked about their justification.

Consideration of the phenomenon of blaming the victim and of our punitive use of certain kinds of expressions of the reactive attitudes shows that a response need not amount to more than a judgment of fault or less than punishment, in order to count as genuine blame, at least as we understand the concept in everyday life. As I tried to show in considering the debate over the practical utility and moral propriety of blaming, many people, philosophers and laypersons alike, are genuinely interested in whether or not we ought to ever blame each other and take it that this question does not have an obvious answer. The theories of Scanlon and Wallace, however, both render the question uninteresting. On Scanlon’s picture, as we saw, to ask whether we are ever justified in blaming others amounts to wondering whether or not we are morally justified in ceasing to confide in people who can’t keep secrets, or in deciding not to let our children spend a night at a friend’s house if an abusive parent lives there. We are clearly justified in modifying our relationships with others in these ways.

Wallace’s approach to blame also distorts the question: For Wallace, that blame be understood as distinct from mere disapproval, sanctioning, and punishment, and that blame have “legitimacy as a form
of interpersonal behavior” are desiderata of the whole enterprise. And while his justification of the reactive attitudes is applicable to the “ought we ever blame?” debate in a way that Scanlon’s discussion is not, his method unnecessarily simplifies the question by restricting its scope in a way that rules out conversation about the appropriateness of disapproval, sanctioning behavior, and punishment, among other reactions we think of as “blame.”

* 

The inability of these accounts to advance our understanding of one of the major questions that occupies those who contemplate the role of blame in our lives is symptomatic of a more general problem I have until now only been gesturing at: Providing an account of blame often serves to obscure rather than to illuminate the blame-related moral questions that interest us. As I said at the outset, this is a problem that arises for all attempts to isolate the “essence” or “special quality” of blame. In the following discussion I focus on three of the most prominent accounts of the nature of blame, but I take my arguments to have a wider scope: I aim to show that this entire literature rests on a mistake.

One of Scanlon’s motivations for offering his account of blame is to provide a convincing response to the phenomenon of moral luck. Scanlon takes it as a datum that in cases of bad “moral outcome luck” involving two people with the same intentions and attitudes, we blame the unlucky agent more than his lucky counterpart. Scanlon sees his task as explaining why this is the case. He writes that it is quite natural for us to think that the “causal outcome of [the unlucky agent’s] action multiplies the significance of his fault,” and that “an adequate theory of blame should either explain how blame can vary in [this] way […] , or give a convincing explanation of why it should appear to even though it does not.”

13 Ibid, 348.

14 Scanlon, 150.

15 Ibid, 126.
Not surprisingly, Scanlon’s account of moral luck explains the difference in blame in terms of the justified changes those affected by the bad outcome can make in their relationships with the unlucky agent. Much could be said about moral luck in general and Scanlon’s treatment of it in particular. My main point is that Scanlon’s insistence on providing a relationship-based account of blame first and then using it to explain moral luck renders the phenomenon implausibly uninteresting and unproblematic. Yes, the parents of a child who was killed by a drunk driver are justified in arranging their lives so that they do not bump into him on the street, but this was never at issue. The interesting problem of moral luck isn’t that we use the words “more blameworthy” to describe the unlucky agent’s situation, but rather that we resent him more, think he should feel guiltier, become angrier with him, complain more loudly about his conduct, and wish to punish him more harshly, than we do a lucky agent with the same attitudes and intentions. When we worry about moral luck, we want to know which, if any, of these practices and others, are justified, and why. Knowing exactly which of these responses to the unlucky agent officially count as “blaming” him is a secondary question at best.

As it turns out, this mistake is pervasive in the blame literature. The theorist begins with a question of general interest that arises in ordinary language (“Should we ever blame people?” “Is it right to blame the unlucky driver more than the lucky driver?”). She then reasons that in order to fully answer the question we must clarify the term “blame.” But rather than asking what the speakers who used the term in raising the initial questions must have meant by it, the theorist undertakes a philosophical investigation that results in a theory of the “nature,” “essence,” or “special character” of blame. But when we plug in the neatly defined concept, the questions we began with seem distorted and uninteresting. The theorizing has done more harm than good.

George Sher frames his book In Praise of Blame around just such an intuitively compelling question. In response to the work of Derk Pereboom, who advocates the rejection of blame, Sher contends that a world without blame would be almost inhuman:

16 Ibid, 151.
The fact that we find it disorienting to imagine a world without blame […] confirm[s] the impression that blame is somehow central to our actual moral lives. Further confirmation, if any is needed, can be found in the extraordinary lengths to which we are willing to go to affix and avoid blame. Many a negotiation, on issues ranging from divorce settlements to the fate of international refugees, has broken down precisely over who is to blame. Reflecting on these facts, we are naturally led to wonder *why* blame is as central to our lives as it evidently is.\textsuperscript{17}

This is the question Sher sets out to answer. Before he does so, however, he must answer, as he puts it, “the surprisingly unexplored question of what blame itself is.”\textsuperscript{18} Ultimately, Sher offers a theory: According to him, blame is “the familiar belief that the person in question has acted badly or has a bad character, but also, second, a corresponding desire that that person *not* have acted badly or *not* have a bad character.”\textsuperscript{19}

Of course, Sher acknowledges, the reactive attitudes are often connected to these belief-desire pairs:

Each blame-constituting desire-belief pair standardly gives rise to a characteristic set of behavioral and affective dispositions. Because the relation between the desire-belief pairs and the dispositions is merely contingent, it would not be inconsistent for someone to acknowledge both the moral importance and the unavoidability of the desire-belief pairs but to deny either the moral importance or the unavoidability of the dispositions to which they standardly give rise. Thus, despite the fact that being fully committed to morality requires that we have blame-constituting desires whenever we have the corresponding beliefs, we cannot automatically infer that it also requires the rancor and recriminations with which blame is commonly associated.

I will not attempt to reconstruct Sher’s argument for this conclusion here, but I will make two basic points about the account. First, I am skeptical of the claim that blaming need involve the desire that the wrongdoer not have acted wrongly or had a bad character. When the opponents of a politician mired in scandal go on national television and gleefully scold him for his misdeeds, are they not *really* blaming him?

But second, and more important, the Sher account makes the crucial mistake we saw in our examination of Scanlon’s treatment of moral luck. He begins with questions about our ability to lead recognizably human lives without blame and develops an account that is supposed to help to clarify and

\textsuperscript{17} Sher, 6.
\textsuperscript{18} Ibid, 6.
\textsuperscript{19} Ibid, 14.
answer these questions. But the theory, it seems to me, only allows Sher to avoid the most compelling issues. By defining blame as the belief that someone acted wrongly or has bad character and the desire that the person not have acted badly or had bad character, Sher has radically changed the sense of the ordinary question he started with. The initial question about the desirability or possibility of a world without blame was perhaps ambiguous, but it seems to me likely that when people argue that we would be better off without blame, they do not usually mean that the world would be a better place if we stopped believing that people were morally bad and wishing they were better. What’s the harm in that? Rather, it seems likely that the people who raise these questions about abandoning blame are wondering about life without the angry attitudes.

Thus, Sher has made things too easy for himself. Ultimately, his defense of blame is not based on consideration of the propriety of anger and resentment, but rather on the importance of our ability to make any moral judgments at all. The cost of abandoning blame, he writes:

would consist not of any diminution of our happiness or prosperity, but rather of our falling in a category that is disturbingly close to that of the wrongdoers and bad people whom we would refuse to blame. Like the person who does not even act on whichever moral principles are justified, the person who does act on those principles but does not care if their demands have been flouted or ignored in the past, or if anyone is disposed to flout or ignore them now, is *ipso facto* not as committed to them as their justification warrants. Even if his failure to form blame-constituting desires has no impact on the quality of his experience, there will remain a gap between the way he lives and the way he has reason to live.20

For Sher, then, to give up on blaming would amount almost to giving up on morality completely, but this thesis isn’t as interesting or surprising as it seems. After all, given Sher’s definition of the concept, to abandon blame would just mean either to stop making wrongness judgments altogether, or to somehow stop caring about them.

Perhaps Sher has answered one possible variant of the “can we live in a world without blame?” question. And maybe those who were curious about the possibility of a life free from wrongness judgments or from the desire that people act morally will feel satisfied by his arguments. But I suspect that some who wondered about a blame-free world will be puzzled by Sher’s conclusion: When they plug

20 Ibid, 134.
Sher’s account of blame into his original question, it is no longer recognizable… they may even lose interest in answering it altogether.

* 

I suspect that another source of the philosophical urge to give a systematic account of blame stems from the acceptance of the thesis that to be morally responsible for an action is to be a legitimate object of blame, credit, or praise for that action. It is a question about whether the truth of determinism has any effect on the ways in which we are justified in treating others. As we have seen, attempting to answer this question by considering the legitimacy of Scanlonian blame is not helpful (determined or not, no one thinks we are not justified in ceasing to associate with people who treat us badly, for example). But the scope of the question “is blame justified?” is not exhausted by consideration of our reactive attitudes either. In this context, the debate between compatibilists and incompatibilists is a dispute in which the truth of determinism puts the legitimacy of anger, guilt, and resentment, but also punishment, moral disapproval, complaint, and, yes, even certain sorts of Scanlonian relationship change, up for discussion and critical evaluation. Perhaps determinism makes no difference to the propriety of any of these reactions. Perhaps it makes a great deal of difference to some or all.

Nomy Arpaly has attempted to diffuse the debate between compatibilists and incompatibilists by arguing that our everyday judgments about blameworthiness, reasons-responsiveness, and moral obligation do not rest on the assumption that we live in a non-deterministic world. 

Like Sher and Scanlon, then, Arpaly begins with an interesting question about blame that could arise in ordinary discourse: “Can anyone be blameworthy if determinism is true?” In order to answer it, she attempts a theory of blameworthiness and blame, offering an account similar to Sher’s, on which to blame someone is to make a judgment with a certain level of investment. Arpaly writes that even though we sometimes speak as if blaming is something we can do to someone else, “This is a category mistake.”

---


22 Ibid, 9.
blame not an action,” she writes, “blame is also not an emotion, nor a feeling, nor any set of emotions and feelings. One can blame Julius Caesar for some of his actions without ever feeling anger, resentment, or indignation toward him.”

For Arpaly, blame is to be understood in terms of blameworthiness, which is the fundamental category: “A person is praiseworthy for taking a morally right course of action out of good will and blameworthy for taking a morally wrong course of action out of lack of good will or out of ill will.” To actually blame someone, one must make the judgment that the target of blame acted with ill will or a lack of good will, and have some investment in that judgment.

The blamer or admirer must not only recognize ill or good will, respectively; he must disapprove of the former and approve of the latter—both in general and in relation to the agent and action toward which the blame is directed.... The moral emotions, while not amounting to praise and blame, naturally stem from them, because they are the natural responses of a person who cares about morality and who perceives good will or ill will in others.

One problem with Arpaly’s account of blame is that its focus on ill will seems to lead to wrong verdicts in certain instances, even if she is right that blame is, “at its core,” a judgment. The family of cases that highlight the point most clearly involve good-hearted people who fail in roles that call for a certain degree of competence.

But once again, there is the larger issue that I have been focusing on: Why say that blame is only a judgment? As we saw, Arpaly’s version of the claim is quite strong, as she goes so far as to say that to speak of blame as if it were an action or feeling is to make a “category mistake.” But what are we to make of this contention, and what is its status? It surely must be a thesis about our concept. What sort of evidence can Arpaly provide for it?

Of course, in many cases, when we say that we blame someone, we are reporting a judgment. But what are we to make of the numerous examples of people who use the term to describe an action or a

---

23 Ibid, 12.
24 Ibid, 15.
feeling? Consider: “I still blame Cousin Ben for ruining Thanksgiving last year!” and let us assume that it is widely acknowledged within the family that Cousin Ben did indeed, through an act of ill will, ruin Thanksgiving last year. What might the speaker mean? On Arpaly’s view, the speaker means that she still thinks that Cousin Ben acted with ill will. It’s true, of course, that the speaker maintains that belief, but the belief doesn’t explain the sentence. If it did, her interlocutor would say something like: “Yeah, so what? We all know what he did.” Isn’t the better explanation of the utterance that the blamer is still angry, still harboring bad feelings, about what happened at Thanksgiving last year? Or consider the truism that forgiveness sometimes involves letting go of blame. Do people who say this mean that forgiveness involves abandoning the judgment that the wrongdoer acted with ill will? No—they mean, roughly, that when we forgive, we get over our anger and resentment.

I think that these considerations show that Arpaly is wrong to hold that ordinary speakers who talk of blame in this way are all making a category mistake, but the more important question question is this: given the prominence of these supposedly mistaken uses of the word “blame,” how can Arpaly be sure that the kind of blameworthiness that we think might be threatened by determinism is the one she has in mind? Were people who worried about the propriety of blameworthiness in a deterministic world really worried about their ability to make quality of will judgments? Perhaps some of them were, but it seems likely that, category mistake or not, the kind of blameworthiness that many people worry about in this context has to do with the legitimacy of our feeling and expressing negative reactive attitudes. Thus, at best, Arpaly’s discussion of blame is limited: if successful, it shows that even in a deterministic world, we can judge that people act well or poorly, and we can have some level of investment in those judgments. I suspect, though, that this modest thesis falls well short even of Arpaly’s own ambitions. Like Scanlon and Sher, Arpaly has, by offering a theory of blame, made the interesting question she started with almost unrecognizable.

The point is the same one we saw in the case of moral luck: to restrict the class of responses we reevaluate in light of the truth of determinism only to the reactive attitudes or only to the holding of certain desires, beliefs, or desire/belief pairs, in order to be faithful to some strictly defined notion of
blame, would be to do ourselves a disservice. We might be interested in the status of any or all of these reactions and perhaps more when we ask if the truth of determinism renders blame inappropriate, irrational, or morally wrong.

* 

Someone convinced by the preceding arguments might now think that we ought to do away with talk of blame in philosophy, as it is imprecise and apt to cause confusion. Wouldn’t we be better off, a reader might wonder, if we asked, one by one, whether resentment, or punishment, or disapproval, or guilt, for example, was appropriate in a given case? Perhaps. Still, when left alone in its complicated position in ordinary language, blame talk can be useful to philosophers. When we ask questions like: “Is it justifiable to blame the unlucky driver more than the lucky driver?” or “Are we justified in blaming people for their actions if determinism is true?” we invite debates with wide scope. Here, the word “blame” functions as a catchall term that prompts us to think about the appropriate reactions to a given case. Of course, in order to make progress in the ensuing moral discussions, we will need to be clearer about exactly what is at stake. Usually, we will need to make more specific claims like “Yelling at him would be excessive, but I hope he feels guilty about what he did,” or “It’s fine for her to express disapproval, but resentment seems too severe.”

* 

Over the course of the next four chapters, I will attempt a systematic undertaking in this spirit, focusing on the feeling and expression of anger and what I will call “reactive” blame. My goals will be to explain and vindicate the social practices that govern and structure these attitudes. In chapter two, I will specify the range of attitudes I am interested in, and then provide a new argument for the conclusion that angry blame is a form of interpersonal interaction we ought to preserve in ethical life. In chapters three and four, I discuss, and reject, the two most prominent putative conditions of reactive blameworthiness, control and self-disclosure. Finally, in chapter five, I propose and defend a practice-based theory of the conditions of moral responsibility and blameworthiness that avoids, and helps to explain, the pitfalls of the control and self-disclosure approaches.
I: INTRODUCTION

Recent years have seen the emergence of a robust literature on the nature of blame, its characteristic features, and its role in moral life. Some of the greatest contributions to this body of scholarship have stressed blame’s function as a means of “moral protest” and its ability to prompt “moral conversation,” or “exchange.” These views have powerfully articulated the aims of successful blaming encounters.

But it seems to me that these theories, though suggestive and important, are limited: on their own, they are insufficient to vindicate our blaming practices. This is because they overlook, and sometimes even distort, the very aspect of blame that stands in greatest need of justification. A fully satisfying account of blame’s role in moral life requires an honest look at the nasty feelings and raw emotional outbursts that led philosophers to question the moral justification or fairness of blame in the first place. Crucially, this examination must go beyond an exposition of what blaming attitudes and reactions aim to accomplish. Indeed, the analysis must expose the characteristic ways in which blame accomplishes its tasks. What, if anything, is distinctive about full-blooded, painful, expressions of anger and resentment that makes them better “conversation starters” or means of moral protest than sadness, disappointment, or stone-cold reasoning? Only after understanding the ways in which anger operates can we begin to

---


29 To her credit, McGeer also sees the limitations of theories that “sanitize” our blaming practices. I argue in section IV, however, that her treatment of anger is itself too “sanitized.”
consider its justification. In this paper, I argue that anger is able to do its work because it is scary; its connection to action and (sometimes violent) threat allows those who employ it to stand up for themselves, establish or reestablish social standing and self-respect, and bring transgressors back into the moral fold.

On the way to this result, this paper explores two related questions about angry blame. First, what does it respond to? Second, what are its aims? After surveying some recent contributions to the blame and resentment literature, I conclude in section II that anger is responsive to threats to things we care about, and especially to perceived slights or what Martha Nussbaum calls “status injuries.” In section III, I argue that anger characteristically aims to right imbalances of relative standing. Drawing from the work of Jeffrey Murphy, Pamela Hieronymi, and others, I defend this goal of status-restoration and respond to Nussbaum’s recent suggestion that the desire to “lower” offenders in the name of status restoration is objectionably narcissistic.

In section IV, I argue that during the recent blame literature boom, philosophers have made substantial progress toward understanding the causes and goals of anger, but have either overlooked or mishandled the question of why anger is especially apt for the moral work we assign it. I claim that by over-rationalizing the moral “conversation,” “exchange,” or “protest” that follows transgression, recent blame theory has left us without an adequate explanation of why anger and resentment are more fitting responses to certain kinds of wrongdoing than sadness, disappointment, or other, calmer attitudes. In other words, it has left us without a fully satisfying justification of anger.

I attempt to fill in these gaps by acknowledging the “ugly” aspects of angry blame: its connection to threat, its tendency to jolt, and its ability to scare. On my view, angry blame, when moderated by the practices that govern its expression, is a valuable part of our moral practice because its association with action allows us to signal our willingness to fight: perhaps for respect, perhaps to maintain or restore a relationship, perhaps on behalf of someone else.
II: WHAT MAKES US ANGRY?

Before proceeding, I should be clear about what I mean when I refer to “angry blame” or just “blame.” We use the word “blame” in various contexts, to describe a wide range of reactions. In this paper, I focus on the expression of negative reactive attitudes, primarily anger and resentment, but also indignation, and other close relatives. Susan Wolf describes the attitudes I have in mind as:

A range that includes resentment, indignation, guilt, and righteous anger – they are emotional attitudes that involve negative feelings toward a person, arising from the belief or impression that the person has behaved badly toward oneself or to a member (or members) of a community about which one cares and which tend to give rise to or perhaps even include a desire to scold or punish the person for his bad behavior.  

As my argument progresses, I will focus especially on angry expressions of resentment, outbursts of righteous anger on the basis of something you have done to me. I have chosen to concentrate on this sort of blaming interaction for two reasons—first, because it seems like a paradigm case of anger, and, second, because, its attendant hostility and antagonism make it seem especially in need of moral justification.

Before turning to that task, of offering a moral justification for the overall practice of expressing anger, I shall first, briefly, consider its internal norms. When is anger fitting? What is angry blame responsive to? One plausible suggestion, offered in various forms by R. Jay Wallace, Christopher Evan Franklin, and Victoria McGeer, among others, is that anger is responsive to damage or threats to things we value and can help us protect those things from harm. Franklin, for example, writes: “blame is a response to free disvaluations of objects of moral value,” and McGeer explains: “At the most elemental level, blame is a quasi-autonomous, emotionally mediated response to others that is specifically prompted


31 See, for example, the recent discussions of blame offered by R. Jay Wallace, Christopher Evan Franklin, and Victoria McGeer in Blame, ed. D. Justin Coates and N.A. Tognazzini. New York: Oxford University Press (2013).

32 Franklin, 217.
by, and targeted on, behavior that transgresses personally or socially valued norms.”

More broadly, Shaun Nichols has written that “moral anger is triggered by perceived injustices.”

But philosophers and self-reflective non-philosophers alike have long noted the tight connection between resentment and a desire to strike back at offenders, and the language of protection and defense is not particularly helpful in making sense of this relationship. After all, one can defend and protect something without causing anyone pain. And so while the defense of valued people, institutions, and relationships is important to our blaming behavior, I think we should, if possible, seek an explanation of angry feelings that improves our understanding of why anger and resentment so often manifest themselves in retaliatory desires and behavior.

Martha Nussbaum has recently offered one explanation of this kind, arguing that anger is (often) responsive to relative down-ranking or “status injury.” Nussbaum, as we will see shortly, thinks that such status-anger is rooted in an objectionable narcissism and is thus morally problematic. But plenty of pro-resentment theorists share her conviction that anger, resentment, and angry blame are responsive to perceived slights or status injuries. Jeffrey Murphy, for example, writes, “I am, in short, suggesting that the primary value defended by the passion of resentment is self-respect, that proper self-respect is essentially tied to the passion of resentment, and that a person who does not resent moral injuries done to him…is almost necessarily a person lacking in self-respect.” Pamela Hieronymi makes similar remarks: “If asked why we resent someone or something or some state of affairs, we typically answer with an

---

33 McGeer, 172.


35 Nietzsche and Joseph Butler are classic examples.


explanation of how that person, event, or state of affairs disrespects us, belittles us, or otherwise threatens our worth....”

Of course, the contention that anger and resentment respond to status injuries is compatible with the thesis that angry blame is directed toward threats or damage to things we value, provided that we value our own status. But there is something deep and right behind the claim that slights and disrespect that threaten one’s standing are especially prone to generate angry attitudes.

Consideration of the ritual of apology and forgiveness can help to highlight the strength of the connection between resentment and status. In offering an apology and asking for forgiveness, the transgressor humbles himself to the wronged parties and (re)affirms his commitment to shared standards. Murphy goes a step further, suggesting that the apology ritual can amount to a kind of humiliation:

Wrongdoers attempt (sometimes successfully) to degrade or insult us; to bring us low; to say “I am on high while you are down there below.” As a result, we in a real sense lose face when done a moral injury—one reason why easy forgiveness tends to compromise self-esteem. But our moral relations provide a ritual whereby the wrongdoer can symbolically bring himself low (or raise us up—I am not sure which metaphor best captures the point)—in other words, the humbling ritual of apology, the language of which is often that of begging for forgiveness. The posture of begging is not very exalted, of course, and thus some symbolic equality—necessary if forgiveness is to proceed consistently with self-respect—is now present.

To be clear, I am not suggesting that every instance of anger or blame is a response to a perceived down-ranking. Still, I think one must focus especially on these kinds of slights to fully understand our angry-blaming practices. Moreover, it seems to me that those instances of resentment that stem from perceived status injuries are most in need of moral defense: some bristle at the thought that we ought to care about relative status at all, and especially at the idea that we should direct unpleasant attitudes toward others to correct imbalances in rank.

---

38 Hieronymi, 547.
39 Murphy, 28.
40 I discuss Nussbaum’s objection to the focus on status in the next section.
III: WHAT DOES ANGER SEEK?

The discussion of apology and humbling leads nicely to consideration of another aspect of angry blame’s internal norms: what is the goal of anger? What does resentment seek? Not surprisingly, theorists who think that angry blame is responsive to threats to valuable things argue that its goal is to protect those things from harm. Franklin and Wallace both hold versions of this thesis, arguing that a failure to respond with resentment or anger can betray an objectionable apathy toward important goods.

I will return to these valuing theories later, but for now I wish to focus on the status-injury accounts. What do philosophers who understand resentment as a response to threats to status or relative standing see as the goal of the angry attitudes? For Hieronymi, resentment is a kind of “moral protest,” a way for the aggrieved to register that he or she will not stand for being treated as less-than: “Resentment affirms what the act denies—its wrongness and the victim's worth. And so, in a way, resentment is a fight response. It fights the meaning of the past event, affirming its wrongness and the moral significance of the victim and the wrongdoer.” For Hieronymi, when one makes an adequate, humbling apology, he disavows the bad action and announces a change of heart. When the aggrieved party forgives the offender, she “ratifies” his change of heart and allows both parties to leave the event “in the past.”

For Nussbaum, anger aims more directly at abasement than protest. On her account, when resentment and anger are rational, they involve the desire to put the offender down, to lower his or her relative status in order to correct the imbalance in rank the transgression caused or exposed. Nussbaum finds this desire to be narcissistic because it implies that the aggrieved party sees the immoral action as being all about his or her status rather than the intrinsic badness of the act itself.

41 Hieronymi, 547.
42 Ibid, 550.
43 Nussbaum, 26.
I’ll make three replies to Nussbaum’s charge of narcissism. First, I am skeptical that the desire to lower offenders need be narcissistic at all. It strikes me that the thought motivating the urge to diminish transgressors may be more along the lines of “no one should be above the law” or “the norms of this community apply to him, too,” as opposed to “he doesn’t get to be above me.” Imagine, for example, feeling resentful of an acquaintance who has wounded you with a mean, insensitive remark, and let’s stipulate that this isn’t the first time he’s made a comment of this kind. It seems perfectly intelligible to me, and very much in the spirit of resentment, that the thought behind your anger might be: “I wish someone would teach him a lesson; he needs to learn that the rules of decency and kindness apply to everyone.” This sentiment, though still involving a desire to lower the offender, is not narcissistic; that is, it doesn’t imply an undue focus on your status, specifically. Rather, it centers on a desire for fairness and equal adherence to the relevant norms.

My second reply focuses on Nussbaum’s claim that a concern with status implies a “tendency to see everything that happens as about oneself and one’s own rank.” She emphasizes that an anger that aims to right imbalances in status “converts all injuries into problems of relative position, thus making the world revolve around the desire of vulnerable selves for domination and control” (my emphasis). Why put the claim in such extreme terms? After all, part of many moral wrongs lies in their ability to diminish the status of their victims and those who care about them. The thought that one’s relative rank has been compromised, or that one has been disrespected, is compatible with, and perhaps in some cases even essential to, a complete understanding of an action’s badness.

Third, and perhaps most important, being concerned about status, even when this concern manifests itself in a resentful desire to “right the balance” of relative rank, is permissible, and in some cases even admirable. Consider the following passage from Murphy, for example, who attempts to vindicate a healthy status anxiety:

44 Ibid, 28.
Intentional wrongdoing *insults* us and attempts (sometimes successfully) to *degrade* us—and thus it involves a kind of injury that is not merely tangible and sensible. It is moral injury, and we care about such injuries…. Most of us tend to care about what others (at least *some* others, some significant group whose good opinion we value) think about us—how much they think we matter. Our self-respect is *social* in at least this sense, and it is simply part of the human condition that we are weak and vulnerable in these ways. And thus when we are treated with contempt by others it attacks us in profound and deeply threatening ways.  

As much as we may not want to admit it, the actions of others *can* diminish us. We can be disrespected and socially lowered. In a certain mood, it is easy to say that we ought not care about our social standing, that we should render ourselves invulnerable to the contempt and status-related disrespect of others by reminding ourselves of our intrinsic worth, moral status, and inalienable dignity. In practice, however, there are times when our dignity really *is* at risk, when we are genuinely “dissed.” Of course there is a sense in which we all have moral worth no matter how others slight us, but we are not wrong to take steps to ensure that our de facto status is equal to others’ in our communities and that we are seen as genuine peers.

To be fair, Nussbaum does acknowledge that some slights *do* amount to affronts on dignity. Using the example of discrimination, she explains that even in these cases, lowering the offender does not accomplish the goal of securing the respect and dignity the mistreated party desires:

> Discrimination on grounds of race or gender, is often imagined as an injury that really does consist in down-ranking, so there is a tendency to think it can be rectified by bringing the injurer low. But this idea is a false lure. What is wanted…is equal respect for human dignity. What is wrong with discrimination is its denial of equality, as well as its many harms to well-being and opportunity. Reversing positions through down-ranking does not create equality. It just substitutes one inequality for another. \(^{47}\)

It is worth noting that many kinds of wrongs can impinge upon one’s dignity or social standing. Even minor slights, such as being interrupted or ignored, can reveal disrespect. Nussbaum is right to say that when others threaten and compromise our dignity, what we want is to be treated with respect. But she does not consider the possibility that this respect and dignity may be partially constituted by the ability to make our resentment felt and understood, and by the ability to prompt wrongdoers to reconsider their

\(^{46}\) Murphy, 25.

\(^{47}\) Nussbaum, 29.
actions and humble themselves in a way that rights the imbalance in status and reaffirms a shared commitment to dignity, equality, respect, or other relevant norms.

This last point, that the process of expressing resentment and provoking an offender to consider his actions and humble *himself* via apology, is worth exploring in some detail. It is not enough to simply say that anger seeks to lower the offender, because, as we saw in our brief consideration of apology, not just any lowering will do. In ideal cases, expression of resentment is not successful unless it results in a kind of *self-humbling* in which the transgressor comes to understand the pain he has caused and the fault in his deed, and asks for forgiveness.\(^48\)

It should come as no surprise, given her criticisms of anger, that Nussbaum finds the apology/forgiveness ritual morally problematic. She writes that “the forgiveness process between humans also focuses unduly on status, suggesting that lowness and abasement compensate for a lowering or status-offense that the offender has inflicted…. The victim is encouraged to enjoy the spectacle of [the wrongdoer’s] groveling as an intrinsically valuable part of the forgiveness process.”\(^49\)

Nussbaum is right, as we have seen, to stress the importance of status in moral exchanges that proceed from anger to forgiveness. Nevertheless, her emphasis on *lowering*, specifically, is slightly misleading. Yes, apology and forgiveness are status-leveling mechanisms, but, as Murphy notes, rebalancing and restoration of status may be achieved either by lowering the offender or by raising the victim up (or both).\(^50\) The point is important because it reminds us that even when we *do* wish for wrongdoers to lower themselves by offering an apology, the goal is not quite for the abasement to “compensate” for the wrongdoing, nor is it to provide an enjoyable spectacle. Rather, we hope offenders will apologize and ask for forgiveness because in doing so they demonstrate their commitment to the

\(^{48}\) McGeer is insightful on this point. For her, the purpose and chief salutary feature of blame is its ability to prompt reflection and response: “On the account I offer, the point and power of blame is to draw wrongdoers into a kind of exchange where they are perforce challenged to exercise their capacities as responsible agents, to reflect on what they have done, whether or not it is legitimate, and if it is not, to *take* responsibility for what they have done and for what they will do in the future” (180).

\(^{49}\) Nussbaum, 74.

\(^{50}\) Murphy, 28.
relationship and its (violated) governing norms. Status-leveling apologies, either through lowering, raising, or both, allow wrongdoer and wronged to proceed on equal moral footing—they offer an assurance that the miscreant does not see himself as “above the law.”

In fact, it isn’t clear apology would even be recognizable without its status-altering features; sincerely uttering “I’m sorry,” “I hope you can forgive me,” or even “You were right and I was wrong,” is likely to force the speaker to “swallow his pride” and to inspire a welcome sense of validation and worth in the recipient. I think this analysis takes much of the sting out of Nussbaum’s concern that in the Christian forgiveness ritual from which our contemporary, secular, apology and forgiveness practices have emerged, “the drama of lowness and fear has been amped up so high that there seems to be no room for personal dignity or self-respect.” What is a transgressor, feeling regret at having hurt someone he cares for, and, perhaps even desperately, wanting to reassure the victim that he sees her as a moral equal, supposed to do? Of course, excessive groveling is undignified, but surely we can leave room for a distinction between bad apologies that involve sickening self-abasement and good ones that involve a self-humbling that is both healthy and tasteful.

IV: SCARY, GOOD

In the last section, I argued that the pain involved in being lowered in terms of relative status, or of lowering oneself through feelings of remorse and guilt, does not pose a threat to the moral justification of the negative reactive attitudes. But feeling another person’s anger involves another kind of discomfort that philosophers have noticed and discussed, and yet have failed to fully face. This is the pain that comes from the characteristic volume and venom of expressions of resentment. Without addressing this aspect of angry blame we will not be able to produce an entirely satisfying justification of the practice.

51 My grandmother requires that anyone shown to be wrong in family arguments—which are understood to range from the most trifling factual disagreements (whether the movie starts at 7:00 or 7:05, for example), to full-blown moral disputes—perform this restorative allocution.

52 Nussbaum, 74. In mentioning fear, Nussbaum is onto something significant, as we will see in the next section.
There are two levels at which one might question the justification or the appropriateness of anger. First, one may look to see whether or not a given outburst of anger is fitting, whether it makes sense given the norms of anger expression. I have suggested that expressions of anger are apt in this sense when one’s status has been unjustly diminished, for example. But we may also consider the justification of anger at another level. Perhaps there are moral reasons against ever expressing anger. Perhaps the emotion, given its jarring nature, is one we ought to deemphasize in our lives, to jettison to the extent that we are able. This is the level at which I will attempt to justify anger in the following sections.

In order to fully understand angry blame one must face exactly what makes it distinct from other emotions, and what makes it distinctly unpleasant; one must honestly reckon with the features of anger that lead questions of its justification to arise in the first place. Next, one must explain why the emotion, and its expression, are respectable (or good) parts of our moral practice, preferably in a way that vindicates the very features of blame that make it appear so morally questionable. These are the main tasks of this section.

V: ANGER AMONG INTIMATES:

One way to see what makes blame unique as a means of moral protest, conversation-starting, or social leveling is to examine alternative means by which we might stand up for ourselves and criticize objectionable conduct. McGeer attempts such an examination and argues that only emotional expressions of our displeasure will do:

Emotions are what biologists call ‘expensive signals’: they are hard to generate under conditions that do not naturally prime a person to experience them; equally they are hard to fake. Blaming emotions thus have a special communicative (and evidential) power. Since they are hard to generate under conditions wherein blamers do not genuinely perceive an injury, they convey to transgressors both how seriously blamers regard the offense and their sincerity in pressing their normative demands.53

I agree that the angry emotions have, as McGeer suggests, a “special communicative power,” but McGeer’s explanation of this power, though helpful, is incomplete. I will argue that what is being

53 Ibid, 182.
communicated is not exactly the seriousness with which one regards the offense, or the sincerity of the moral objection, but rather another kind of seriousness: a willingness to fight.

Before I introduce this alternate proposal, let us consider McGeer’s suggestion that emotions convey the sincerity of a normative demand. Consider the following case: in a moment of frustration, you belittled a close friend. Now imagine that the next time you see this friend, he pulls you aside and, in a firm, emotionless tone, explains to you that what you said to him hurt him deeply and that if you do not produce the apology he feels you owe him he will no longer spend time with you. To my mind, an angry expression of these thoughts, as opposed to your friend’s measured response, would not indicate that he somehow takes the offense more seriously, nor would it add sincerity to his moral demand. In fact, in many cases, a lack of emotion conveys an even greater seriousness than an expression of anger. If someone is ready to listen, emotionless explanation of the meaning and consequences of the transgression can signal grave seriousness and utter sincerity.

Moreover, even if emotions are sometimes necessary to convey a sense of seriousness and sincerity, McGeer’s argument (which focused on the difficulty of faking emotions and the energy required to generate them) is not specific to the angry and blaming emotions.

McGeer may agree with these points. Her thesis was not that anger is always necessary to convey seriousness and sincerity. I do not mean so much to object as to provide what I take to be an important elaboration. My complaint is not that McGeer is wrong to say that the angry attitudes can communicate seriousness and sincerity. Rather, my worry is that her account has not yet isolated the distinctive features of anger that allow it to do its characteristic work. A full defense of resentment and its kindred attitudes would explain what is special about them, as opposed to, say, sadness and disappointment. It would show what makes these responses so well-suited to the task of getting people to rethink their actions.

Both Wallace and Franklin make remarks that suggest an answer to the question of the unique power of the angry emotions. Here is Wallace:

Vulnerability to this emotional response represents a way of caring about moral values that uniquely answers to their relational character…. People who care about this form of relationship typically hold themselves and others to the moral norms that are constitutive of it, where this in
turn involves a susceptibility to the distinctively reactive sentiments. When we are wronged by another, we are not just saddened by their failure to relate to us on a basis of mutual regard; we resent such treatment, and this emotion has a relational aspect that fits the character of moral norms as constituting a valuable form of human relationship. To resent someone is to feel not merely that they have acted wrongly, but that they have wronged us in particular, violating the norms that constitute relations of mutual regard.54

According to this analysis, resentment and indignation are relational, especially apt for protesting violations of my dignity or your rights. Sadness and disappointment, on the other hand, are not capable of responding to the relational aspect of moral values in this way.

Franklin also offers analysis that has the potential to explain the special power of the blaming emotions. He criticizes Wallace for failing to explain why the appropriateness of blame for responding to one’s having been wronged is enough to justify the practice: “Although blame is a way of valuing the values at the heart of morality, Wallace does not show that it is essential to valuing these values, but only that it is one way among others (my emphasis).”55 Nevertheless, his explanation of anger’s distinctive power ends up looking similar to Wallace’s with its emphasis on the “relational” aspect of anger: its ability to respond to someone in particular having transgressed against or diminished something the victim values.

Two features of sadness—what it responds to and what it expresses—prevent it from playing the same role as blame. Sadness is fundamentally a response to the loss of or harm to an object we value, whereas blame is a response to the object’s being freely disvalued [by an agent]. We respond with sadness when an object we care about is lost or harmed, regardless of the cause of the harm…. Sadness only tracks the fact that loss or harm occurred…, and so cannot protect and defend moral values in the ways identified above. Sadness also, and most important, does not have the dimension of condemnation required for defending and protecting moral values. This is because sadness does not indicate that anyone has violated the standards of value. It indicates only that an object of value has been harmed, but not all harms are disvaluations.”56

But it seems to me that although generic sadness or melancholy may not reflect the kind of commitment to value Wallace and Franklin have in mind, certain kinds of sadness and disappointment do so quite well. First, consider an example of sadness in a close relationship: If I decide to blow off my

54 Wallace, 369.

55 Franklin, 213.

56 Franklin, 221-2.
sister’s housewarming party, her sorrow will not take the form of some abstract melancholy, but rather
sadness that I, her own brother, would bail on her celebration. It seems implausible to hold that this sort
of sad reaction indicates that she somehow cares less about my feelings toward her than she could, or that
she sees my absence as anything less than a serious “disvaluation” of her and her preferences.

To take another example, I am sure I am not the only one who, as a child, dreaded disappointing
my parents more than angering them. When I disappointed them, they were not disappointed that
someone had done a bad thing; they were disappointed in me because of what I did.57

Certain kinds of disappointment and sadness, then, express genuine concern for valuable objects
and relationships, but a sad or disappointed reaction as opposed to an angry one can reveal a great deal
about a respondent’s interpretation of a slight and his or her relationship with the transgressor. The
meanings of these different reactions, as well as the reasons why sad, disappointed, and angry responses
can take on these meanings, have been underexplored.

The problem, I think, is that philosophers have tried to locate the force of anger and resentment in
too “rational” a place. My suggestion is that anger confronts us on a more “basic” level than a “rational”
presentation of reasons, arguments, and even feelings. Expression of anger is effective in getting
transgressors to reconsider the meaning of their behavior, and prompting them to engage in the status-
balancing apology ritual, because of its connection to action, even violent and physical threats. Simply
put, anger is scary. It is because of this, perhaps often unconscious, association with threat and danger that
expressions of anger and resentment have their attention-grabbing power.

When wrongdoing occurs in the context of a close relationship, interpersonal disappointment and
sadness can, of course, be appropriate, but anger has an importantly distinct place in these intimate
contexts. I have been arguing that expressing disappointment or sadness to someone, like expressing
resentment, can be a way of conveying one’s (often very deep) disapproval. But disappointment goes
beyond resentment by revealing a kind of despair: rather than fighting to keep the wrongdoer in the fold,

57 I do not mean to suggest that “interpersonal” or “directed” sadness and disappointment always arise in response to
a transgression. They do not. One can be disappointed in one’s child for his choice of college major or in a friend for
her bad taste in music, for example.
disappointment can signal withdrawal and reevaluation.\textsuperscript{58} This is why it can sting so deeply to learn that one has disappointed or saddened a close friend or family member: if the disappointment is justified, then one has failed to live up to the standards that govern the relationship, failed to be or become the kind of person the wronged party thought one was or could be.

This kind of serious shortcoming is in general harder to undo than a transgression that would trigger anger. Anger does not necessarily imply reevaluation of character or weakening of the relationship. In fact, in many cases it implies the opposite: namely, that the offender, through humbling himself and apologizing, can more or less set things right and the parties can put the incident behind them. In anger we \textit{fight back}, we fight for the relationship to remain as it is, return to how it was, or realize its potential. Thus, willingness to fight for a relationship rather than reconsidering it or withdrawing from it can be a sign of trust and investment. Expressing anger as opposed to disappointment or sadness allows us to communicate a certain kind of faith, or at least hope, that the episode that triggered the anger will not force us to reassess or terminate the relationship.\textsuperscript{59}

It is no accident that receiving this message, feeling this willingness to fight, is jarring, unpleasant, and sometimes even intimidating. In order to fully understand the connections between anger, payback, and the restoration of status and self-respect, we must face up to anger’s scariness. Expressions of anger are not distinctive because of their ability to convey sincerity or to call attention to specifically interpersonal slights; rather, the crucial general feature of anger is its promise of \textit{action} should the imbalances that triggered the anger persist.

\textsuperscript{58} Of course, distancing oneself from the wrongdoer or reevaluating the relationship does not necessarily imply taking the objective attitude toward her.

\textsuperscript{59} Here I depart again from Nussbaum, who suggests that disappointment, not anger, is generally the more “well-grounded” response to minor wrongdoing: “Some minor wrongs in a marriage may be occasions only for disappointment: they undermine expectations, but they don’t really undermine trust. Thus if one person is always late, that may be annoying, but it won’t feel like betrayal, unless punctuality has assumed unusual importance in the relationship, or unless unpunctuality is plausibly read as a sign of something deeper, such as a lack of respect” (115).
VI: ANGER TOWARD STRANGERS

What about anger toward strangers? Once again, we may begin by considering the effectiveness of expressions of sadness or disappointment in getting miscreants to rethink their actions. Suppose you cry in response to a friend wrongdoing you. If he is truly your friend, and cares about you and your feelings, seeing your tears may be disconcerting to him, and will perhaps force him to consider his remarks and the pain they caused you. He will feel pangs of remorse and guilt and be prompted to apologize. Now imagine being wronged by a stranger. Will seeing you cry affect him at all? Perhaps, depending on who you are and who he is, and what social positions you occupy (sex, race, sexual orientation, class, and a host of other factors could be relevant). But for a particularly hard-hearted stranger, an expression of anger, which will be more likely to jolt, scare, or destabilize, may be required to do the job. Sometimes only a threat will do. The accounts of Hieronymi, Wallace, and Franklin, with their language of protest, protection, and defense, discerningly call attention to this tight connection between anger and action.

In section III, I argued that protests or defenses like these aim to restore equality of status between parties. Anger may lead to this kind of rebalancing in two ways. So far I have been discussing cases in which the offender needs a jostle, but then he sees the error of his ways, renounces his bad action, and offers an apology. This is, to borrow McGeer’s term, a “mentalized” version of the process. I suspect, however, that in many cases, and especially in cases of stranger-directed anger, the language of “conversation” and “exchange” is a misleading over-rationalization of the blaming process. Athletes who play contact sports, for example, often have occasion to make use of short bursts of anger (“Get your elbow out of my face!”). These small blaming encounters are successful when the opponent retreats, stops playing “dirty.” Apologies and feelings of remorse are irrelevant; the important thing is that he backs down. Often, these reactions are not “mentalized” in the way McGeer, Michael McKenna, and others

60 There is, of course, another reason disappointment seems out of place when directed toward strangers: These emotions are most at home in relationships defined in part by high expectations of goodwill between the parties. Just as strangers cannot betray us, strangers, and even acquaintances, cannot disappoint or sadden us like our close friends or family members can.

61 In fact, I take it to be a merit of this analysis that it can explain this close relationship between anger and protest.
have in mind, but this doesn’t automatically unjustify them: the dirty player disrespects his opponent and the norms of the game; in expressing anger and getting him to back off, the angry player reestablishes his standing.

One might be tempted to respond that this sort of case isn’t an instance of genuine blame. But that would be a mistake. We can construct moral cases along the same lines. Imagine a woman responding to a catcall on the street by whipping around, locking eyes with the harasser and yelling: “Don’t you ever talk to me like that!” It seems to me that whether or not the woman’s anger prompts the catcaller to question the moral justification of his behavior or feel remorseful is beside the point. At least in some plausible version of the case, that is not the anger’s aim. Is it thereby unjustified, or somehow undeserved? Quite the opposite: the anger is an apt, and morally justified, response to disrespectful behavior that attacks the victim’s de facto social standing; it forcefully communicates a readiness to fight for respect.

The catcaller case is obviously moral, but is it blame? The question is a bit odd. Examples like these clearly involve righteous anger: one has been wronged and lashes out in protest, asserting his or her status. These are precisely the sorts of outbursts that led philosophers to wonder about the justification of angry blame in the first place.

These examples serve to highlight another feature of anger that disappointment and sadness do not share, and this is that there is a boldness to anger that other responses to wrongdoing lack. Anger often triggers anger in return; in expressing it, the victim risks potential backlash. But one’s willingness to provoke is itself a sign of strength (“I don’t care if I get my ass kicked, I’m going to speak my mind!”), and demonstrating such courage can facilitate self-respect and the restoration of status.

Of course, not all expressions of anger are provocative in this way. Marilyn Frye has observed, for example, that a woman’s anger is often taken less seriously than a man’s. Frye focuses on social expectations of women and their ability to occupy positions of authority: she notes that is harder for women to have their anger and resentment taken seriously in the workplace than in the kitchen, for example. It is worth elaborating on Frye’s thought in a way that extends the position I have been

developing here about the (sometimes unconscious) connection between anger and physical violence: one reason a man’s anger is taken seriously is because failure to take it seriously can be dangerous. Moreover, men, and people in positions of power more generally, can safely express anger in situations in which less powerful others cannot. It is safe for a rich, white American to yell at his immigrant housekeeper, but not the other way around. My account helps to make sense of these power dynamics.

Do these imbalances in the ways agents may successfully employ anger threaten its justification? No, although acknowledging the tight connection between anger and threat can help us understand some of the moral boundaries we must (and do) draw around the expression of anger, and why those limitations are highly contextual, varying even in accordance with the physical stature of the angry person.

In practice, no group has a monopoly on the effective and justified use of anger. This is in part because anger’s scariness need not always involve an explicit or implicit threat of physical violence. In many cases of anger, especially among intimates, the “scariness” of anger could take the form of a threat of a temporary withdrawal, or a temporary withdrawal of goodwill. These sorts of threats are nonviolent, but they still harness the characteristic jolting power of anger by suggesting a willingness to make the target’s life more difficult if a change in behavior, an apology, or a restoration of status is not forthcoming. Consider, for example, the phenomenon of giving a family member “the silent treatment”.

The aggrieved party gives the wrongdoer a threatening glimpse of what a more permanent withdrawal would look like, and the “treatment” is effective when the vision is unpleasant or jarring enough for the transgressor to apologize or make a show of contrition. The “scariness” in this case comes in the form of a silent withdrawal, but one that expresses anger rather than disappointment. The wrongdoer is invited to reestablish communication and continue the relationship on terms the silent victim can accept. A more permanent version of the “silent treatment,” a total termination of the relationship, may be experienced as jolting to the transgressor as well, but the interaction is different. The goal of the withdrawal in that case is the withdrawal itself, not scaring the wrongdoer straight; there is no invitation to go back to the way things were.

63 Susan Wolf suggested this example.
Still, we should not diminish, nor should we totally disavow, the relationship between anger and physical violence. In fact, in certain cases, it is the very threat of violence that makes anger an effective tool for putting a stop to degrading or bullying behavior without coming to blows. The questions of when and how we may employ, or threaten, violence in order to right or to prevent a moral wrong are notoriously difficult, and the answers will vary depending on the specifics of the cases at issue. I cannot address these questions here. Provisionally, I will say that the kinds of threats I have sought to justify here must at least be accompanied by an intelligible (and legitimate) moral demand or complaint that addresses the target in a common moral language, based upon shared, or at least socially accepted, principles. Anger, even when threatening, directed toward strangers, and seeking no remorse or apology, must retain a distinctively human moral character, even if is not explicitly “mentalizing” in McGeer’s sense of the term.

I also do not mean to suggest that all anger is always scary, or that it is only justified when it is scary. In this paper, I have focused on angry, “in-your-face” expressions of resentment, first, because such outbursts seem to me to be centrally important to our understanding of the negative reactive attitudes, and, second, because these emotional surges have seemed, to both philosophers and laypersons, particularly scary and unpleasant, and thus especially in need of justification. But that is not to say that this sort of anger is the only kind of anger, or the only important or morally interesting kind. It is no surprise that my account will not perfectly capture other species of the emotion: One may, for example, become angry with her favorite team’s quarterback when he throws an interception, and never intend, or threaten, to do anything about it. And then there is anger on behalf of others, a phenomenon more closely related to the kind of angry blame I have concentrated on here, but that I have not addressed in detail. There are important analogies, and disanalogies, between resentment and these varieties of anger. I think my remarks will provide at least a start toward helping us better understand the purpose and method of those angry feelings, but they cannot tell the whole story on their own. Here, my central claim was only that angry expressions of resentment toward intimates and toward strangers are characteristically scary. And while this connection to threat and action may manifest itself in different ways depending on the
relationships between the parties in question, it is the scariness and jolting power of angry blame that distinguishes it from other negative emotions and makes it an effective tool for restoring social status and self-respect.

Finally, I should emphasize that I am not claiming that anger is only morally permissible when it is the only way to get the target’s attention in a given case. As I wrote earlier, I am concerned with justifying expressions of angry blame as a general practice. The appropriateness of anger in individual cases will hinge on the considerations of fittingness I addressed briefly in sections II and III.

VII: CONCLUSION: TWO PLEAS FOR COMPOSURE

I take the forgoing remarks to provide a moral justification of angry blame. Life without the angry attitudes, in addition to being alien, would be impoverished, depriving us of the ability to fight for our relationships and respect with dignity and authenticity.

But perhaps the anger skeptic is still unconvinced. “Isn’t the idea behind angry blame, that people need a visceral jolt to change or reconsider their behavior, a bit cynical?” we can imagine him asking. “Aren’t we giving up on good old rational conversation too easily?”

Responding to this objection is instructive because it brings out a false dichotomy implicit in the contemporary blame literature: on the prevailing model, either blame is “mentalizing,” to use McGeer’s term, or it is objectionably primitive; either it is a part of a rationalized “responsibility exchange,” or it is immoral bullying, manipulation via non-rational threat. One of the aims of this paper is to carve out a middle ground between these extremes. I have argued that in order to fully understand the power of blame we must acknowledge (and even endorse) its connection to threatening action. It is naïve to think that every rational agent will be instantly receptive to reasons and argument. It is often difficult, inconvenient, and embarrassing to back down, to reconsider one’s actions, or to allow oneself to feel guilt and remorse. Even caring, responsible people can be stubborn and dense, needing a “non-rational” push to engage in these painful processes. The point could be read as an extension of Bernard Williams’s discussion of
“proleptic” blame: sometimes signaling a willingness to stand up and fight for one’s respect, or for the restoration of a relationship, gives the blamed agent a reason to care. ⁶⁴

Moreover, it is worth noting that the phenomenon of “pushing” someone by engaging him at a more “basic” level is not exclusive to blaming. Conveying one’s joy at the prospect of an activity, or offering rewards for good behavior are also forms of “non-rational” motivation, as are displays of sadness, disappointment and countless other emotional expressions. ⁶⁵

Warming up to the possibility that some level of coercion and non-rational threat might be essential to an effective blaming practice, though, may raise another kind of question. One might wonder why emotion was necessary at all; wouldn’t a system in which we dispassionately stated our demands and our threats (“Unless you apologize to me, I will hit you, shame you, withdraw good will, etc.”) be more honest and straightforward?

There may be many reasons to prefer our angry blaming practice to this dispassionate alternative; here I will present two that I see as especially important. First, emotional responses signal a kind of vulnerability that stone-cold communication of facts does not. When we lash out in anger, we reveal that our aggressor’s actions can hurt and threaten us. One might, for example, implore an angry friend engaged in a dispute with an antagonist not to “give him the satisfaction” of a reaction. I submit that the satisfaction in question is status-related: it is, roughly, the satisfaction of having gotten under the target’s skin, of knowing that she sees her social rank or self-respect as unstable enough that some status-restoration is necessary in response to the offense. ⁶⁶

---


⁶⁵ The notion that the expression of any emotion could be objectionably manipulative raises another important question: namely, to what extent are these emotions, and our expressions of them, under our control? This is a question I cannot answer here. In this paper, I have discussed the purpose of a certain kind of anger and the way it functions in moral life. I have attempted to isolate patterns in our blaming behavior and make some arguments that will, at least, bear upon us when we can exercise control over our emotional response.

⁶⁶ Thus while the expression of resentment can often facilitate self-respect, in other cases, a lack of response, or a contemptuous reaction, can be empowering, indicating that one cannot be wounded by the aggressor’s disrespectful behavior. Sorting out exactly which of these reactions are best suited to which cases is important ethical work that I
Angry reactions, then, precisely because they involve a loss of control that betrays vulnerability, have a meaning, and a conversational role, that is importantly distinct from their bloodless alternatives, which are more appropriate for interactions with children or others who pose little threat to our social standing (“If you don’t apologize, I’ll send you to your room.”). Where expressions of anger suggest emotional exposure and equality of status, composed statements of demands and threats can signal reserve, detachment, and even superiority. These postures are not only incongruous with the ways in which we experience moral life, they are often outright objectionable in the context of close relationships in which we are, and ought to be, emotionally vulnerable. Thus, calmly and straightforwardly stating that one plans to make an aggressor’s life worse in some way if he or she doesn’t apologize would not serve the same social function as an angry reaction even if the linguistic contents of the two responses were identical.

But even the assumption of identical content is dubious, and this is my second response to the cold-blooded proposal. The implicit threat that attends an angry reaction will often be much less specific and content-rich than a calm and clear presentation of demands and penalties. Straightforwardness is not always a virtue; an angry expression of a moral demand allows a blamer to jolt her target without explicitly articulating a stakes-raising threat.

This makes expressions of anger “conversational,” in a way that has perhaps been underappreciated. Calm, specific, articulation of demands and threats is limiting, both for the blamer and the blamed: It constricts the blamed agent’s possible responses, and, in turn, forces the blamer into a corner if the blamed agent calls the bluff or doubles down on his bad behavior. In this way, a reactive system that sanctioned only the composed approach, like other practices that truncate conversation in favor of an explicit statement of policy and consequences, could foster imperiousness and arrogance, and could unduly restrict one’s ability to protest. Angry outbursts, on the other hand, allow victims to leave

 cannot undertake here. For now, it is enough to say that we need room in the moral toolbox for all of them. For more on contempt as a salutary moral attitude, see Bell, M. (2013). *Hard Feelings*. Oxford: Oxford University Press.
interchanges open-ended because the threats implicit in their blaming responses are less than fully determinate.

* 

Still, a blamer must respect his audience; obviously we often go overboard with anger. This is especially true for people in positions of power for whom expressing anger carries minimal risk, and who can intimidate their targets easily. Of course we should be careful not to overdo it with expressions of resentment, but this is hardly news. My point here is only that we must leave a place for angry blame in moral life; we should neither banish it nor squint until it appears defanged, scary though it is.

\[67\]

As I have said, what exactly counts as overdoing it is also a matter for consideration, though one that I cannot address here. In one especially interesting class of cases, blamers might continue to yell and scold even when the offender has lowered him or herself and offered a fulsome apology. This sort of blaming strikes me as more purely punitive (though not necessarily wrong or inappropriate) than many of the responses I’ve considered here.
I: INTRODUCTION

Many philosophers of moral responsibility and blame, both compatibilist and incompatibilist, are control freaks, who see it as almost axiomatic that blameworthy action must be subject to some degree of agential control.

These theorists typically introduce the control condition on moral responsibility and blameworthiness by prompting their audience to recognize the familiarity of “I had no control over it” or “I couldn’t help it” as valid excuses for otherwise objectionable conduct, and then generalizing from those cases to a sweeping control requirement. Fischer and Ravizza, for example, use Aristotle’s discussion of force as an excusing factor to introduce some classic cases in which a blamed agent would offer such a defense:

As paradigm instances of forced actions, Aristotle offers examples in which a person is carried somewhere by the wind against his will, or taken somewhere by kidnappers who have him in their power. To this pair of cases might be added other stock examples in which responsibility is undermined by force: such cases might involve irresistible psychological impulses, brainwashing, hypnosis, or direct manipulation of the brain.68

In these cases, it plainly seems wrong to direct negative reactive attitudes toward the agents in question, and their lack of control is a plausible explanation for their blamelessness.

How we are to understand the control condition, however, is not always clear. Some theorists seem to see it as a conceptual truth. Others see the principle as an ethical one, arguing that blame for the uncontrolled is unfair. Whatever the status of the requirement, many of us share the intuition that blame is not, in normal circumstances, justifiably directed toward agents who lacked control—that “I couldn’t help it” is often a legitimate excuse.

In this chapter, I offer one of the strongest rejections of the control condition on responsibility and blame in the contemporary literature, and one that I hope can help to explain the requirement’s initial plausibility. I argue that even strong forms of reactive blame, full-blooded expressions of anger and resentment, can be morally justified in response to actions over which the blamed agent had neither direct nor indirect control. To make the case, I survey the ways in which philosophers have justified the control requirement. First, I consider the possibility that control and blameworthiness are conceptually linked. After examining the ethics of angry blame in response to behavior that is neither directly nor indirectly controlled, I reject this proposal. Second, I consider the thesis that angry blame for uncontrolled behavior is unfair or undeserved. I argue that this suggestion mischaracterizes the fairness and desert we ought to aim for in reacting to wrongdoing: blame is fair when it is sanctioned by a justified blaming practice, and these practices do not, and indeed should not, provide a perfect match between degrees of negative reactive feedback and degrees of controlled wrongdoing. Building on the arguments of the previous chapter about the value of blame and the way it works, I suggest that angry blame is responsive to disrespect and other threats to social standing that are highly contingent and contextual and do not require control to produce.

II: THE CONCEPTUAL INTERPRETATION

What would it mean for the control condition on blameworthiness to be conceptual? The thought must be something like this: If one fully understood the concept of blameworthiness, one would see that it required control. One way to undermine this kind of alleged conceptual relationship is by counterexample; in this section I will introduce and discuss a set of cases that cast doubt on the assumption that our competent rulings of blameworthiness imply a judgment that the blamed agent had control over his action. The examples show that despite the truth of the blamed agents’ insistence that
their actions were out of their control in whatever way the control theorist requires, we do not, and should not, excuse the offenders from blame.⁶⁹

Before introducing the cases, I should note that any plausible version of the control requirement must include some kind of “tracing” provision. Consider a familiar kind of drunk driver case: in the moments right before the driver strikes the pedestrian, he may be doing everything he can to avoid the collision. Still, it is obvious that citing a lack of control would not excuse him from blame. In general, control theorists accommodate these sorts of cases by invoking a “tracing” procedure: if the bad action (or even emotional state) can be “traced” back to some “root” action subject to the agent’s control, then the agent may be blamed for his subsequent act. Of course, causal chains are long and complicated, and so not just any action can serve as the root of the trace. Fischer and Ravizza, for example, hold that for a tracing procedure to successfully explain an agent’s moral responsibility, “it must be reasonable to expect the agent to know that his prior exercise of guidance control will or might well lead to his subsequent emotional state or reaction.”⁷⁰

The drunk driver case is one classic illustration, but the tracing strategy should be familiar from all sorts of legal and interpersonal examples. Tracing captures the thought that one isn’t excused from blame and punishment if he recklessly puts himself into a dangerous situation in which he loses direct control over his actions. In such cases, “I couldn’t help it” is not a valid excuse because, the blamer would explain, the agent could have avoided being in that situation in the first place. The tracing strategy, then, is essential to any control-based theory of blameworthiness: without tracing, a theory with a strict control condition would yield the wrong verdict in cases in which the agent is intuitively blameworthy for his actions despite their not being under his direct rational or voluntary control.

---

⁶⁹ Of course, there are many ways of specifying the kind and degree of control required for responsibility and blameworthiness. I have omitted discussion of the specifics of these accounts because my argument is meant to be quite general and in dialogue with any plausible theory of responsibility that includes a control condition. The cases are not meant to succeed because of a technicality or an idiosyncrasy of any particular view.

⁷⁰Ibid, 257.
Having clarified the role of tracing, we can now turn to the cases. It’s a Friday night, and some college roommates are happily watching TV in their off-campus apartment when they hear a knock at their door. They open it, welcoming two of their friends, Al and Bill. Al and Bill have clearly been drinking, but Al is in much better shape. Bill is “blackout” drunk, and possibly in danger of alcohol poisoning. Bill staggers into the room. Seeing his audience, a crazed glint enters his eyes. He begins ranting lustfully about one of the roommates’ girlfriends, describing horrifying, unprintable, acts. The roommates tell him to stop, but this only provokes him. He turns his attention to another one of his friends and hatefully unleashes a storm of racial slurs and insults.

My argument that Bill is blameworthy is based on my impression of the ways in which the roommates in the story would be justified in responding to his ugly rant. I contend that they would be within their rights to feel and express disgust and resentment toward Bill, and ask him to leave their apartment. Imagine they do so, and that when Bill awakes the next morning, he has no idea what happened the night before. The last thing he remembers is merrily taking shots of tequila at a party. When Al fills him in on the night’s events, Bill is mortified. He feels ashamed, but more than that, he feels guilty. In addition to embarrassing himself, he has wronged people he cares about. He apologizes, making a show of contrition and begging for forgiveness.

Now consider the next morning’s developments from the perspective of the roommates. They are justified, it seems to me, in feeling hurt, disgusted, and, perhaps most significantly for my purposes here, angry. This is especially true for those Bill singled out during his rant. They are also justified in looking at Bill with a kind of moral shock and horror that, while different from anger, is not the sort of attitude one could properly take toward a child or someone with a serious mental illness. In other words, it would be permissible, and even good, for them to feel and express a range of reactive attitudes during their confrontation with Bill. Indeed, one can imagine a morally virtuous person, and exemplary friend, feeling and expressing these sentiments. Expressions of anger and resentment, in addition to being fitting, might be especially good: angry and resentful reactions could help Bill’s victims to reestablish their social standing and foster self-respect.
I suspect that most would agree with these observations. Those skeptical of them should imagine Bill explaining to his friends that although he agrees the incident was unfortunate, he doesn’t feel guilty or remorseful because he was too intoxicated to have any idea what he was saying, much less rationally control himself. The roommates, it seems to me, if they ever wished to interact with Bill on friendly terms again, might reasonably try to prompt Bill to feel guilty about his behavior. To do so, they would explain and express their anger, disappointment, shock, and pain. Even if Bill is contrite and apologetic, the roommates might reasonably think less of Bill as a person, or note a demerit on his moral record. Once it is agreed that all, or even most, of these reactions would be appropriate, the question of Bill’s blameworthiness seems settled.

Obviously, Bill lacks direct control over his actions, and so to argue for his blameworthiness, the control theorist would turn to a tracing strategy. But tracing in this situation is misplaced. Is Bill’s diatribe reasonably foreseeable at any point? Suppose Bill had no idea that the racist, sexist thoughts he expressed on the fateful night were even present in his psyche at all: how could he be expected to foresee his terrible outburst? Perhaps the control theorist would argue that even if taking tequila shots is not morally wrong, it is risky; any college student should know that when he or she drinks too much, bad things can happen. I wish to make two points in response to this modification. First, this move would widen the scope of the trace beyond plausibility. Almost every action we take has risks, and a tracing model that interprets “reasonable foreseeability” such that Blackout Bill’s rant counts as reasonably foreseeable simply because drinking tequila is risky seems to stretch the notion of “reasonable foreseeability” beyond recognition, especially if we assume that Bill had no idea the racist, sexist thoughts were buried in his mind.

Second, and more importantly, the judgment that Bill is blameworthy would not be confused even if we could not trace his reprehensible behavior to any plausible root action. Imagine, for example, that Blackout Bill is inexperienced with drinking and, through no fault of his own, doesn’t “know his limits.” Or take the most extreme case and imagine that Bill was forced to drink.\textsuperscript{71} It seems to me that even in

\begin{footnotesize}
\begin{enumerate}
\item This version of the case is importantly different from a similar case one might imagine in which Bill is hypnotized and compelled recite offensive lines fed to him by the hypnotist. In the versions of the case I have in mind, it is
\end{enumerate}
\end{footnotesize}
these versions of the example, blaming reactions to Bill’s behavior would be coherent, and this is because in Bill’s case we lack even the *instinct* to trace. We often cannot know whether or not the agent, at some prior time, could have reasonably concluded that his or her controlled action might result in some future bad act, but, more importantly, we *don’t care.*

Of course, the control theorist is not yet out of tracing options. She might suggest that our pre-theoretical verdict that Bill is blameworthy is based on the assumption that he had control over his *becoming the sort of person who would think or say these terrible things in the first place.* Again, I’m not convinced that any such assumption is involved in the commonsense blameworthiness judgment. It seems to me that the roommates could accept Bill’s insistence that his words do not truly reflect his stable character. They might have no reason to think that Bill actually endorses, or has ever even indulged, any of the terrible thoughts he expressed in his rant. In fact, one might even think that all Bill’s tirade proves is that he has successfully, perhaps even admirably, *repressed* these thoughts in the sober, rationally controlled part of his life. But even this angelic version of Bill should feel guilty about his actions and apologize to his friends, who would still be justified in feeling angry, disgusted, and betrayed.

Michael McKenna has endorsed a character-based tracing strategy of this kind. He argues that, “a unifying requirement on moral responsibility is that control comes in somewhere. […] In the case of direct moral responsibility for the nonvoluntary, [control] comes in indirectly, via a (sometimes

significant that the thoughts are Bill’s own despite the fact that he does not rationally endorse them and may not even be aware of them. A hypnotist case that is closer in spirit to the example I focus on here (though still importantly different) would be one in which the hypnotist compels Bill to reveal some of his deepest, perhaps even suppressed, desires or feelings, leading to horrifying results.


73 Of course, the fact that apology is appropriate does not always mean that the agent is blameworthy for the action. It is appropriate, and good, to apologize for, say, faultlessly colliding with someone while walking on the sidewalk. I have something stronger in mind here: in this case, I am suggesting that the apology is *owed* because Blackout Bill has *wronged* his friends.

74 Labeling himself a “control freak,” as it happens.
unexercised) capacity to decide freely to evaluate one's moral standpoint (as regards the relevant nonvoluntary object of responsibility).”

Thus, to make the case that Bill was blameworthy, McKenna could hold that in quieter, sober times, Bill possessed “a standing capacity to perform a free mental act of deciding or choosing to evaluate [his] moral standpoint(s),” and this capacity, even if unexercised, constitutes sufficient control to ground Bill’s blameworthiness.

Again, though, the modification will not solve the problem. Perhaps no reasonable amount of self-reflection and evaluation would have ever revealed to Bill that he was capable of erupting in the way he did. And even if Bill had realized that he had the potential to explode, this realization on its own would not be enough to ground his later blameworthiness. Suppose he was taking the reasonable steps to reform himself; this fact would not undermine the coherence of the blaming reaction.

The story of Blackout Bill is one of a large family of problem cases for control freaks in which blameworthy agents lack control and in which we have no impulse to trace. Consider, for example, an instance of road rage in which a driver is overwhelmed by frustration and anger and takes a 5-iron to another motorist’s car. Or imagine a sleep-deprived parent who beats his child: it seems to me that the child would be well within her rights to feel resentful and angry, and that attempting to trace the sleep deprivation back to some earlier failure of the parent to, say, get to bed at a reasonable hour, would be absurd. Even if one was tempted to engage in a tracing process, it wouldn’t help: perhaps the parent had to stay up late to work or was kept up all night by the sounds of traffic or screaming children. We may


76 Ibid, 36.

77 David Shoemaker, in a recent response to McKenna, has proposed another possible “root” one could trace to: a failure to exercise “empathic control,” in order to get oneself into the right “perceptive stance” by “robustly taking up the affected agent’s normative perspective.” I think the case of Blackout Bill and the other cases I discuss put pressure on this sort of control as well. It seems to me that one could lack empathic control (because of drunkenness, sleep deprivation, or rage, for example), or could have done all that is in one’s power to exercise it, and yet still be blameworthy for one’s bad behavior. See: Shoemaker, D. (2015). McKenna’s quality of will. *Criminal Law and Philosophy, 9* (4), 695-708, 705.

78 The following examples are only brief sketches, and depending on how the particular control condition in play is spelled out they may not all apply to every control theorist’s account.
even imagine that the sleep-deprived child abuser and the road rager were both diligently working through an anger management course and doing what one could reasonably ask of them to become calmer people. Would that mean that their victims’ resentment was based on a conceptual mistake? Or consider a “crime of passion”: a husband comes home to find his wife in bed with another man and impulsively kills them both. The husband’s action might be morally better than murder “in cold blood,” but surely he is still the legitimate target of reactive attitudes, and attempting to trace his blameworthiness to some prior action would be odd.

Reflexive actions that involve no conscious choice at all on the part of the blamed agent are also non-tracing cases. A spontaneous laugh at a funeral, for example, might be met with justified anger and is not intuitively a plausible candidate for tracing. True, we can imagine the laughing person saying “I couldn’t help it,” but did the blamers ever doubt that? Presumably, they already knew that the action was not subject to the laughers’ control when they became angry with him.\footnote{Angela Smith raises some case of forgetfulness that might also serve as examples of this phenomenon. For Smith’s explanation of those cases see: Smith, A. (2006). Control, Responsibility, and Moral Assessment. 
Philosophical Studies, 138(3), 367-392. For commentary see: McKenna, M. (2008), and the final section of Fischer, J., & Tognazzini, N. (2009). The Truth about Tracing. Nous, 49 (3), 531-556.} Or, finally, consider the inciting incident of the recent film Force Majeure: a man eating breakfast with his family at a ski resort instinctively runs to save himself from an avalanche, wordlessly abandoning his wife and children. When the avalanche stops just short of the restaurant and the man returns, his wife, I think rightly, blames him for his solo escape attempt.

These examples tell against the thesis that blameworthiness conceptually requires control. These are cases for which it seems odd to even begin to trace the agents’ bad behaviors back to some suitable root action, and in which blamers were never under the impression that the blamed agents enjoyed any sort of control. Thus, to establish the need for a control condition, the control theorist will need another explanation. That “I couldn’t help it” is sometimes excusing (if true) will only get her so far; it does not ground a fully general control condition in cases in which no one would offer or accept that sort of plea as an excuse.
III: THE ETHICAL INTERPRETATION

It might seem odd, and perhaps unsatisfying, if the control condition could be rejected based simply on the plausibility of a few counterexamples. I suspect that this is because many control freaks have something more explicitly ethical in mind. While some theorists who favor a strong control requirement on blameworthiness explain the condition by citing the intuitive legitimacy of lack of control as an excuse, there is often an implied assumption about fairness or moral propriety lurking in the background. How could it be fair or just, one might ask, to subject someone to unpleasant negative reactive sentiments on the basis of that which he could not control?

There is more than one version of the concern that blame for the uncontrolled is undeserved or unfair. One of these centers on the claim that blame must respond to the person being blamed, to his or her character or “self.” When a person suffers from a sudden and unpredictable seizure, for example, his bodily movements tell us nothing about him. Typically, actions we control express something about us—our desires, beliefs, commitments, and values. One might argue that while it can make sense to respond with blaming reactions to this sort of self expression, it would be wrong to blame people for actions that tell us nothing of significance about the agents who perform them.

This is a philosophically legitimate concern, and I will address it in the next chapter. It cannot, however, be the objection at the heart of the control condition. After all, many of the characters I have introduced do express themselves, or at least some of their attitudes and desires, through their uncontrolled actions. We can imagine that Blackout Bill, for example, expresses some deeply held (though perhaps repressed) beliefs and desires in his raving, or that the protagonist from Force Majeure reveals selfishness and a lack of concern for his family when he runs to save himself from the avalanche.

The versions of the claim that blame for the uncontrolled is morally objectionable that actually underlie the control requirement must involve something like the thought that no one morally deserves to be subjected to negative reactive attitudes on the basis of that which she could not control, or that
blameworthiness requires moral wrongdoing, which in turn requires control. This line of argument has some initial plausibility: one might think that in blaming someone we address him with a kind of demand that he not have done what he did, and that this demand seems to imply that he must have had some level of control over his action. To sharpen the point further, one might appeal to the intuition that it is wrong or unfair to blame people who had no reasonable opportunity to avoid becoming the targets of these unpleasant reactions.

These ways of leveling the complaint that blame for the uncontrolled is unfair raise questions about what blame is, what it responds to, and how it works. The first of these, what blame is, is easy enough to address through simple stipulation: by “blame,” I mean what one might call “reactive blame,”: the feeling and expression of negative attitudes like resentment and anger.

The other questions will be more difficult. To fully appreciate their answers, we must first recognize that reactive blaming, like punishment, is a practice governed by rules. Having focused on the

---

80 R. Jay Wallace endorses a view of this kind, arguing, roughly, that justified blame is a fitting response to failures to meet obligations, but that one does not genuinely have an obligation if one cannot meet it, and thus blaming someone who non-culpably lacked control would amount to blaming someone who did nothing wrong, which is unfair. See: Wallace, R.J. (1994). Responsibility and the Moral Sentiments. Cambridge, Mass.: Harvard University Press.

81 George Sher’s Who Knew (See: Sher, G. (2009). Who Knew? Oxford: Oxford University Press) features a nice discussion, and rejection, of this thought (see, especially, chapters four and nine). The control condition is not Sher’s central focus; when he takes up the topic explicitly his concern is to show that denying the control condition is not necessarily a problem for the sophisticated version of the knowledge condition on responsibility he defends throughout the book. I see Sher as an ally against the control freaks, but my arguments differ from his in a few crucial respects. The most important difference stems from Sher’s understanding of blame for an action as, roughly, a belief that the agent in question acted wrongly and a corresponding desire that the agent would not have acted wrongly (for more on Sher’s conception of blame, see Sher, G. (2005). In Praise of Blame. Oxford: Oxford University Press). Leaving questions about the essence of blame aside, I go beyond Sher in arguing that not only may agents be blameworthy in Sher’s sense for uncontrolled behavior, but, that they are often proper targets of specifically angry and resentful reactions on the basis of that behavior. In order to defend this more ambitious thesis, as we will soon see, I employ an entirely different kind of argument, one that understands the expression of these strong reactive attitudes as being governed by a social practice that allows us to fight for social standing and self-respect.


83 Of course, unlike some systems of punishment, the rules governing reactive blaming are social norms rather than formal legal codes.
mechanics of justified blaming practices in chapter two, I will not give a full accounting of those rules here. It will be helpful, though, to provide a quick sketch of our system of negative reactive attitudes that roughly explains what they respond to and how they work in a set of central cases.

As Pamela Hieronymi, Victoria McGeer, and others have suggested, the angry emotions are characteristically responsive to disrespect. As emotions, anger and resentment play a distinctive role in moral “conversations”:

Unlike even highly personal kinds of disappointment and sadness (“I’m disappointed in you because of what you did”), the angry attitudes have a jolting power stemming from their (often unconscious) connection to (sometimes violent) threats and retaliation. As such, they can grab the attention of the blamed agent while affording the blamer a kind of self-respect. I follow Jeffrey Murphy in thinking that expression of these attitudes, with their implicit demands of apology, can be especially apt for correcting imbalances in relative social rank that arise because of the blamed agent’s actions and their consequences.

When expression of angry attitudes is successful, it can, in one sort of ideal, “rationalized” case, prompt the target to reconsider his or her behavior and to feel the pain of guilt and remorse. Feeling and expressing guilt serves to humble the offender and correct imbalances of social standing that resulted from the behavior in question. In other cases, blaming can be successful if the transgressor backs off or changes his behavior, or simply because the blamer is able to express, with dignity and authenticity, her willingness to fight—perhaps for respect, perhaps for the dignity and equality of standing that will allow the blamer to maintain or restore a relationship, or perhaps on behalf of someone else.

---


85 For a defense of this focus on relative status, see: Murphy’s contribution to: Murphy, J., & Hampton, J. (1988). Forgiveness and Mercy. Cambridge: Cambridge University Press, 16.

86 Manuel Vargas takes a “two-level” approach to responsibility that is structurally similar to the one I offer here but that has radically different results (See: Vargas, M. (2013). Building Better Beings. Oxford: Oxford University Press). Vargas is concerned to explain “responsibility practices” generally, rather than our practices of anger expression, specifically. And, in large part because of his control freak-ish contention that responsiveness to reasons
This understanding of our reactive practices emphasizes the ability of the angry attitudes to help right imbalances in de facto social standing, to respond to disrespect that attacks one’s status. But disrespect and de facto social lowering of the victim (or raising of the offender) are highly contingent phenomena, subject to luck and consequence. Blackout Bill’s racist rant, for example, takes on a different “meaning,” or social significance, to use a term from the work of T.M. Scanlon, because of the context of its delivery. The presence of that particular audience, with their specific histories and relationships to Bill, is essential to Bill’s disrespect. The tirade is not just morally wrong, it also generates a permission for the roommates to blame Bill by directing anger and resentment toward him in an effort to protest his actions and, perhaps, attempt to restore a distribution of de facto social status that must obtain in order for Bill to continue to interact with the roommates in the future.

What does this mean for the control theorist who claims that reactive blame is responsive to wrongdoing, and that wrongdoing requires control, or who alleges that blame for the unavoidable is always undeserved? If I am right that in a large and important set of cases, blame is responsive to disrespect and threats to relative social standing, then the control condition would have to amount to something along the following lines: an agent cannot disrespect or diminish anyone unless he has a certain kind of control over his actions. But this is false; the cases I have focused on are counterexamples to such a thesis. The kind of wrongness the control theorist has in mind (a kind of wrongness that requires control) is not what blame is responsive to. Thus, this version of the complaint that it is immoral to blame...

---

87 Of course not every blaming reaction need function in this way. I am only offering an outline that illuminates what I take to be some centrally important cases.

88 Scanlon, 52.
someone for an uncontrolled action erroneously assumes that the negative reactive attitudes are meant to address only a narrow set of bad behaviors; it miscasts blame in an artificially limited role.  

This practice-based understanding of blame and blameworthiness also explains the implausibility of the claim that the harshness of our reactions to bad behavior should perfectly track the blamed agent’s level of controlled wrongdoing, or the poor quality of will expressed in his or her action. It is immediately obvious that even the most innocuous uses of tracing seem to license the expression of harsher reactive attitudes toward unlucky agents than toward their lucky counterparts. In one sense, we might say that two drunk drivers, for example, are “equally blameworthy,” even though one is unlucky and hits a child while the other is lucky and returns home without incident. But when it comes to the responses sanctioned by our moral practice, the character and tone of the reactive blame that is appropriately directed toward the drivers is subject to huge variation depending on what happens on their trips home. As Thomas Nagel puts the point in “Moral Luck”:

> The same degree of culpability or estimability in intention, motive, or concern is compatible with a wide range of judgments, positive or negative, depending on what happened beyond the point of decision. The mens rea which could have existed in the absence of any consequences does not exhaust the grounds of moral judgment. Actual results influence culpability or esteem in a large class of unquestionably ethical cases ranging from negligence through political choice.

Because reactive blameworthiness is inescapably shaped by “outcome” luck, proponents of tracing theories cannot make good on the promise of blame’s absolute or “metaphysical” fairness. Yes, they

---

89 To be clear, in understanding blame in this way one does not commit oneself to the thesis that Bill and the others do not act wrongly. The thought that their actions are not wrong stems from an artificially narrow understanding of wrongness based, perhaps, on a Kantian conception of obligation.

90 On this point, my account differs from quality of will-based views of blameworthiness that also dismiss the control condition. Nomy Arpaly defends perhaps the most prominent view of this kind in the recent literature (for a thorough elaboration and defense of the model, see the first chapter of: Arpaly, N. (2006). *Merit, Meaning, and Human Bondage*. Princeton: Princeton University Press). Arpaly holds, roughly, that to blame an agent is to recognize and disapprove of his or her failure of good will, and that an agent’s level of blameworthiness depends solely on the deficiency in quality of will expressed in his or her action. Because I am discussing harsh, full-blooded expressions of reactive attitudes like anger and resentment rather than judgments of blameworthiness, it is no surprise that my rejection of the control condition must take quite a different form. I am attempting to make sense of angry and resentful reactions to bad behavior that respond not merely to the quality of will expressed, but to the meaning of the action, described in social terms. My approach takes account of quality of will, but only as one factor among many that are relevant to determining an action’s meaning.

require some controlled action, perhaps quite minor and in the distant past, to ground blameworthiness, but they cannot realize the dream of proportionality between degrees of faultiness and the character of the appropriate reactive attitudes. We want to resent, scold, or yell at agents not in response to some prior root actions (or at least not only in response to these root actions) but, rather, to decry the ultimate actions that have produced the negative consequences. We resent the abusive parent because he struck his child, not because he failed to get a good rest the night before a stressful day; we scream at the road rager “You wrecked my car!” For good reason, our blaming practice is not designed to deliver the abstract fairness characterized by perfect proportionality. A system that did would be an entirely different kind of system, one that might not even be recognizable as a social device meant to facilitate effective and authentic response to disrespectful behavior and threats to relative status.

At this point it might be objected that while blaming agents for uncontrolled action might produce good consequences (restoration of social order, reconciliation, etc.), they nevertheless do not deserve to be blamed. It is worth thinking carefully about exactly what is meant by “deserve” here. If I am right about the contours of our blaming practice, then there is a sense in which it is obvious that, say, Blackout Bill deserves to be blamed: the rules of the practice allow for his being blamed, and if that practice is justified, then blaming Bill is justified. Similarly, if the runner is tagged before he reaches base in a game of baseball, he deserves to be called “out.”

What then would it mean to say that a character like Blackout Bill, or the sleep-deprived child abuser does not deserve to incur his victims’ resentment? It seems clear to me that on a commonsense conception of moral desert, Bill and the other characters do deserve the reactive blame the practice calls for. Imagine Blackout Bill is in your house talking about what he’d like to do to your mother! It is not as if he was suddenly possessed by Satan or manipulated by evil scientists—the thoughts came from him, and when you make him aware of your resentment, horror, and disgust, he gets what he deserves.

Is this kind of desert “basic,” in the sense that some recent contributors to the moral responsibility literature understand the term? Explaining the notion of basic desert, Derk Pereboom writes, “In the basic form of desert, someone who has done wrong for bad reasons deserves to be blamed and perhaps
punished just because he has done wrong for those reasons.” Based on this definition, the commonsense notion of desert I am working with is indeed “basic”: blamed agents deserve to be blamed simply because they have acted in a threatening or disrespectful way. The rules of the reactive practice I have sketched give the blamer permission to blame under these circumstances just because of the meaning or significance of the action, even when we have reason to believe that reactive blame in a particular instance might lead to bad consequences on balance. Transgressors deserve to be blamed because their disrespectful or threatening behavior triggers a permission set out by rules of a justified practice.

But perhaps the suggestion is that the practice itself is morally objectionable, that the reactive system as I have described it is unfair because it sanctions harsh treatment (expression of negative reactive attitudes like resentment) toward people who do not really deserve it. Again, it seems to me that this argument is confused. My suggestion was that our blaming practices set out rules that permit us to feel and express negative reactive attitudes when someone has insulted or disrespected us, or otherwise behaved in such a way as to present some threat to the balance of relative standing within a relationship. I have also suggested that one need not have either direct or indirect control in order to engage in such disrespectful, insulting, or otherwise status-disrupting behavior.

Does the rule that permits Bill’s friends, or the abused child, or the wife in Force Majeure to express anger license blame for undeserving recipients? I fail to see how it does. The allegation, as I understand it, seems to trade on a misconception about the kind of fairness we have reason to be concerned about. The commonsense desert that I have been discussing is, as far as I am concerned, all the desert one should demand. In fact, to claim that reactive blame sanctioned by a socially useful practice, which specifies rules for responding to disrespect and denial of social status, is unfair because the blamed agent lacked control in a particular case would be to make a kind of category mistake. Our system of reactive responses helps us to get the attention of and impose costs on offending parties, protest against


objectionable conduct, clarify and reinforce our standards, and stand up for ourselves and the things we value in a way that authentically reflects our feelings. The reactive attitudes, such as friendship, love, trust, and many other morally valuable aspects of human life, do not always answer to the demands of an abstract, metaphysical fairness. Sometimes, after all, life is unfair: It is unfair, in some sense, that not everyone finds his or her soul mate; it is unfair that the most diligent student does not always receive the best grade. 94 This kind of unfairness, though sometimes unfortunate, is not indicative of any moral problem. Similarly, complaining that the child’s resentment of her abusive father, for example, is objectionably unfair because of his lack of control seems to involve a misunderstanding: it is to demand a kind of fairness that morality need not provide.

Of course, this does not mean that fairness is irrelevant to the legitimacy of blame. Rather, it means that we must be careful in considering the kind of fairness we have in mind when evaluating a blaming reaction. A more promising fairness argument might be that it would be unfair, and unjustified, to direct reactive blame toward someone on the basis of an action that was not disrespectful or threatening in any way, for example. This is the sort of fairness claim that can gain traction in response to misguided reaction: Rather than concerning ourselves with abstract or metaphysical notions of fairness and desert (the sense in which it is “unfair” that someone did not find his soul mate), we should guard against the sort of unfairness that obtains when one is not treated in accordance with the practice that governs the norms of response (the sense in which it is unfair for the umpire to call a runner “safe” if he is tagged before he reaches base). 95

Still, control matters, and matters a great deal, to our reactive blaming practices. In all of the cases I have focused on, I argued that control was not required for blameworthiness. Nevertheless, in every case, the meaning or significance of the actions in question would be drastically different if they were under the agents’ control. (It is bad to go on a racist, sexist tirade when blackout drunk, but cold-blooded expression of the same attitudes when sober would be different, and possibly even more

94 I am grateful to Geoff Sayre-McCord for helping me express this point.

95 This line of argument, though promising, is still too simple. I will explain why in the next chapter.
upsetting.) A fully developed account of a justified reactive blaming practice would leave room for tracing as well: that an agent should have reasonably foreseen that some action of his could lead to some awful outcome is relevant to how we ought to feel about him and his relationship to that bad result; and an agent’s having had the chance to avoid a bad situation is morally significant and should change the way we respond to him in light of the outcomes he caused.

Reflection on these facts suggests that control, understood in an everyday sense, comes in degrees and helps to shape appropriate blaming response. A reasonable blaming practice will not entirely dismiss the intuition that there should be some relationship between the blamed agent’s level of controlled wrongdoing, or controlled expression of ill will, and the degree of appropriate angry feedback. Indeed, a justified and well-functioning reactive system will put this thought in its proper place: Controlled actions often reflect one’s values, the quality of one’s will, and a certain degree of reflection and planning; all of these factors are likely to alter the meaning of one’s conduct and thus the proper reactive response.

IV: CONTROL, PRACTICES, AND INCOMPATIBILISM

Until now, I have mostly been discussing compatibilist control theories. But incompatibilists, who hold that no one is morally responsible in a deterministic world, can be control freaks as well. Like her compatibilist counterpart, the incompatibilist control theorist would endorse the principle that one can never be blameworthy for something he or she did not control. Then, once the principle is in hand, she would explain that because the truth of determinism rules out our having the kind of ultimate control she is interested in, we cannot deserve blame for anything we do.

Before I respond to the incompatibilist argument from fairness, I should note that the incompatibilist strategy I have in mind is different from the more “metaphysical” approach that P.F. Strawson was concerned to thwart in “Freedom and Resentment.”

metaphysically incompatible with determinism. In objecting to this metaphysical approach to questions of justified moral reaction, Strawson wrote, “the fact of our natural human commitment to ordinary interpersonal attitudes…is part of the general framework of human life, not something that can come up for review as particular cases can come up for review within this general framework.” The kind of incompatibilist argument I have in mind attempts to challenge responsibility from within the framework of our existing responsibility practice. This incompatibilist argues that there are, in everyday ethical life, rules and principles that, when combined with the truth of determinism, entail that no one is ever blameworthy for anything. The ethical rule that this kind of incompatibilist would appeal to is some version of the principle I alluded to above: that we do not deserve unpleasant treatment (in the form of the negative reactive attitudes) in response to actions and events we have no control over. The argument, then, is that when people blame each other, it is because they are making a mistake: they think that the blamed agents have a kind of ultimate control over their own actions that they actually lack. If blamers appreciated the arguments for determinism, this “internal” incompatibilist alleges, they would be forced to revise their judgments of reactive blameworthiness.

This argument relies on the assumption that we are, in everyday moral interaction, actually committed to the principle that it is unfair to blame a person for actions he does not control. But the reactions to Blackout Bill’s outburst (and to the actions of the other characters I imagine) that I have defended suggest that we are not so committed. The roommates blame Bill for his racist, sexist rambling (and think that they are morally justified in doing so) without ever thinking that he had rational or voluntary control over himself during the diatribe. True, if Bill explains to them the next morning that he

---

97 Of course, the arguments for and against this thesis are various, and I will not discuss them here.

98 Ibid, 91.

99 Derk Pereboom, for example, makes an argument of this kind: “I believe Strawson is wrong to maintain that a theoretical challenge to the reactive attitudes based on the thesis of universal determinism is external to the practice of holding people morally responsible, and therefore illegitimate. Rather, in my view, exemptions from moral responsibility that are widely regarded as acceptable and thus internal to the practice will generalize to an incompatibilist condition on moral responsibility.” (Pereboom, D. (2001). Living without Free Will. Cambridge: Cambridge University Press, 95).
“had no control over what was coming out of his mouth,” they may go easier on him, but even if they accept this remark they could reasonably continue to blame him by, for example, harboring and expressing anger and resentment, demanding an apology, hoping he feels guilty, and thinking less of him as a person.

Cases like this raise an uncomfortable dilemma for the “ethical” incompatibilist: either the principle that it is unfair to blame people for uncontrolled action is not as important to our actual ethical practice as she claims, or we would be morally wrong to blame Blackout Bill, the road rager, the sleep-deprived child abuser, the parent who simply forgets to pick her child up from school, and countless other seemingly blameworthy agents even if we did, in general, possess incompatibilist-approved control.

Of course, the incompatibilist’s position was always going to be revisionary; she was already content to claim that our everyday blaming interactions rest on a mistake. But that was a different kind of revision: there, the suggestion was that because they lacked information about determinism, ordinary people did not recognize an opportunity to apply a principle they already endorsed. But consideration of Blackout Bill and friends reveals that no such principle is widely accepted, and, without the principle, the “internal” argument for incompatibilism collapses.

We are now in a position to return to some remarks I made at the outset about the way philosophers tend to motivate the control condition. There, I cited Fischer and Ravizza, who use examples in which agents are blameless because they lack control in order to infer a general control requirement. As we have just seen, the incompatibilist control theorist makes the same move, but with a different understanding of the kind of control blameworthiness requires. One way of understanding this chapter is as an attempt to expose and explain the pitfalls of this generalization strategy. The generalization fails because lack of control in these examples is not, on its own, a sufficiently deep explanation of the agents’ blamelessness. In order to know whether such a generalization from the blameless cases is warranted, we need to know why the agents’ lack of control renders them unfit objects of reactive response. And this explains the intuitive pull of the control condition: often, a lack of control does undermine blame, but not because of a conceptual, metaphysical, or ethical requirement. Rather, lack of control mitigates blame
because of its affect on the meaning of the offending action and on our understanding of the extent to which the behavior really communicates something threatening or disrespectful. Insofar as it affects the offending action’s social significance, control is relevant to the content and tone of blame that is warranted, but it is not a necessary condition on reactive feedback.

What, then, is the critical necessary condition of responsible and potentially blameworthy agency? Readers who have found the arguments of this chapter compelling may reasonably wonder whether the expression of one’s judgments, quality of will, or “self,” which occurs in many of the cases of uncontrolled behavior I have discussed, can play this role. This is the question I take up in the next chapter.
Responsibility and the Shallow Self\textsuperscript{100}

I: INTRODUCTION

Imagine attending the opening of an art exhibition. Passing by a beautiful sculpture, you decide to smash it on the floor, just for the fun of it. The artist, on hand for the event, is understandably furious and expresses anger, shock, and resentment. These emotions are fitting, but they would have been misplaced had you been pushed into the sculpture, had a mad scientist taken control of you and forced you into the act, or had you knocked it over during a violent, involuntary muscle spasm.

What do pushes, mad scientist interventions, and spasms have in common that standard blameworthy actions lack? One reasonable answer is that in all three of the “blameless” cases, the action is not the agent’s own in some morally significant way. The scientist manipulation and the physical push, for example, involve alien causes in the form of other agents; the spasm involves the agent’s own brain but tells us nothing morally interesting about him.

Now suppose you are defending a friend to an acquaintance. The acquaintance complains that your friend has been acting erratically, and in his defense, you explain that your friend “hasn’t been himself” since his divorce. In offering this bit of information, you intend to make an excuse for your friend: you hope that your acquaintance will suspend or temper his negative reactive response to your friend’s behavior. Consideration of these two examples illuminates what I take to be the central thought behind “deep self”\textsuperscript{101} or self-disclosure views of moral responsibility and blameworthiness: In order to be responsible for an action, the self-disclosure theorist holds, the action must come from the person, his or her self, in some meaningful way. To articulate this sense of action ownership, the deep-self or self-

\textsuperscript{100} A slightly altered version of this chapter is forthcoming in \textit{Philosophical Studies}.

disclosure theorist appeals to the intuition generated by the second example: what separates actions that
come from our deep selves, on the one hand, and actions that our bodies cause but do not really come
from us, on the other hand, is that the former reveal something important about us, reflect our characters,
or, at least, the quality of our wills.

In this paper, I canvas various ways of understanding the deep self. In doing so, I try both to
isolate some coherent notion of the deep self and to discuss its relationship to the robust sort of moral
responsibility associated with reactive responses like blame and praise. In the second section, I conclude
that self-disclosure views get something right: we do care about how, and to what extent, actions reflect
personalities; we are rightly interested in what we can learn about people through their deeds, reactions,
attitudes, and beliefs. Nevertheless, I argue that the connection between the deep self and moral
responsibility is tenuous and resists easy explanation. In the third and fourth sections I sketch and respond
to the most plausible self-disclosure views on offer in the philosophical literature. I conclude that in many
cases responsibility and blameworthiness do not depend on the agent’s actions expressing anything
especially significant or “deep”; “shallow,” non-disclosive actions that reveal nothing about the quality of
an agent’s will can be blameworthy, even in a strong reactive sense. I argue that self-disclosure views
overestimate the extent to which our blaming practices and responsibility judgments are responsive to
agents as opposed to actions, and that this mistake has the potential to distort not only our understanding
of justified blaming practices, but also our conceptions of the “selves” of blamed agents.

II: THE SELF IN EVERYDAY LIFE

Philosophers are not alone in their interest in the self. As a society, we are intensely occupied
with the ideal of true-self expression. But what is the connection between the deep self and moral

---

102 I use “self,” “real self,” and “deep self” interchangeably in this paper. I also use the terms “deep-self views” and
“real-self views” synonymously with “self-disclosure views.” Some theorists who hold, say, quality of will or
judgement-based theories of moral responsibility may object to the characterization of their positions as a “deep-
self” views, but terminology is not my main concern, as long as my point is clear: I mean to object to any model that
holds that responsible and blameworthy action requires the expression or reflection of an agent’s judgments, cares,
or the quality of his or her will. On Angela Smith’s view (See: Smith, Angela. (2005). Responsibility for Attitudes:
responsibility? One can only begin to understand the philosophical import of the deep self after one gets a sense of our everyday interest in it. What is the real self? Why do we care about it?

We invoke the deep self in a variety of situations. The following is a non-exhaustive list, an attempt to put the concept in context. To begin, we might employ deep-self talk when someone’s behavior is substantially at odds with what we have come to expect of her, when she “isn’t herself.” One might think, for instance, that a family member in the late stages of Alzheimer’s “isn’t there.” Some users of drugs prescribed for depression, bipolar disorder, and other psychological problems report feeling that the drugs make them feel “like themselves,” while others express the opposite sentiment, professing that the drugs make them feel like different people entirely. We usually take these reports to be definitive.

Not so with a substantial number of cases of recreational drug use in which we also render judgments of loss of self. Whereas it seems clear that dementia, for example, obscures or erases the afflicted person’s true self, our judgments about the deep self and recreational drug use are substantially cloudier. “In vino veritas” is, after all, a well-known saying: we often think that drinking, and some kinds of drug use, can actually reveal, rather than obscure, a person’s true character. But this is not always true. When a drinker or drug user does something unfortunate, we sometimes make excuses for him, especially if he is inexperienced with the substance: “he wasn’t thinking straight…, that wasn’t really him,” “that’s not the guy I know,” “that was the liquor talking.”

Instances of stress, sleep deprivation, and fear can function similarly. On the one hand, there is the thought that trying conditions can rob us of our agency. As the Snickers’ tagline tells us: “You’re not you when you’re hungry.” On the other hand, we sometimes say that stress brings out our deepest or most significant traits. Legendary UCLA basketball coach John Wooden held that “Sports don’t build character. They reveal it,” expressing a sentiment echoed in the classic war movie trope in which the Activity and Passivity in Mental Life. *Ethics 115*(2), 236-271), for example, morally responsible action must be reflective of an “evaluative judgment,” but such judgments can be “one-off,” “out of character,” and need not be revelatory of an agent’s deepest commitments. For Smith, we are only responsible for actions that reflect evaluative judgments or commitments and that involve the sort of rational activity one could be called upon to justify. This is “deep” enough for my purposes here: I will argue that morally responsible behavior can fall short even of this standard, that some blameworthy actions reveal *nothing* of significance about their authors. I am grateful to an anonymous reviewer from *Philosophical Studies* for urging me to clarify this point.
A jaded sergeant dramatically tells the recruits, “you don’t really know what you’re made of until you’re out there in some bunker, with only your brothers, far from Mommy, surrounded by the enemy….”

Interestingly, this feeling is compatible with the excusing power of stress: we sometimes judge that stress prompted the expression of a deep and important part of the agent’s personality and yet conclude that his responsibility for the resulting action, and certainly his reactive blameworthiness for it, is diminished.

Another set of cases in which we employ the idea of the deep self involves the familiar appeal to an agent’s “true” commitments and character. A parent might think that her son really does care about being successful deep down but is just lazy; one might warn a friend that a job candidate “is on his best behavior now, but when he gets comfortable his real personality will come out.”

This sort of deep-self talk, implying that only certain actions express our deep moral cores, is related to another use of deep-self discourse in which one appeals to a person’s circumstances or upbringing in order to explain or excuse his or her behavior. Like the stress case, this kind of deep-self talk can lead in opposite directions. On the one hand, the idea that there is a soul or moral center inside of everyone, but that circumstantial factors prevent it (the real person) from shining through, is a familiar one. On the other hand, we often hold that a man is “a product of his circumstances,” or that “his upbringing made him the man he is today.” Despite their opposition on the question of the nature of the real self, both of these reactions to a rough upbringing can play a role in excusing bad conduct.

Next, there is the injunction to “be oneself,” and the related process of “finding oneself.” Once one has “found himself,” by organically accumulating a set of values and projects, we might encourage him to be himself when he is about to meet someone new or enter a stressful environment; job interviews, first dates, and situations involving peer pressure come immediately to mind. Sometimes, we cannot “be ourselves” around people who are hostile and judgmental, and we appreciate the ways in which our close friends and family allow us to be who we truly are.

Let us now turn more directly toward the relationship between responsibility, “out of character” action, and expression of the self. Imagine a waiter who is normally painstakingly polite finally snapping at the end of a long day of work. The diners unlucky enough to be in the restaurant for his eruption might
think they have seen his “real self.” But depending on what the waiter says and does, he might turn out to be blameworthy regardless of whether or not he actually revealed anything especially deep or important about his character. If he really was just having a bad day, for example, and the incident was a genuine aberration, he could still be a legitimate target of the reactive attitudes. The general point is that many blameworthy actions are not revelatory of anything persistent and stable in the agent’s character, and this is reflected in our ordinary understanding of the deep self.

Of course, this survey of our interest in the deep self it is not exhaustive, but I think it is sufficient to situate the concept in its everyday context. What can be gleaned from this overview? Obviously, we can see that the real or deep self is important to us. The sheer diversity of cases above indicates that we are, across a wide array of settings, deeply concerned with real-self expression.

But perhaps the cases are so diverse that they resist systematization. Is there just one notion of the deep self underlying all of this use? Certain examples suggest fragmentation. Recall the case of stress: some say that stressful situations obscure the “real” self, while others claim such situations reveal it. In fact, I am optimistic that both camps are, at least in most cases, out to capture and describe the same thing. At the very least, a vast majority of these examples suggest some notion of a person’s core—a stable, enduring, personality.

The relationship between this important everyday conception of the deep self, however, and judgments of responsibility and blameworthiness, is less clear. It seems to me that much of our interest in the deep self is concerned not with questions of agency, or the fittingness of praise and blame for actions, but rather with notions of identity and authenticity. Consider again the command to “be yourself,” and imagine that a man on a first date becomes nervous and, as a result, ignores the injunction and puts on an elaborate facade of macho posturing, embarrassing himself and insulting his date. His failure to be himself is morally interesting, but this does not imply that his responsibility for his boorish behavior is somehow mitigated. Even on the assumption that his actions really were out of character and explicable in part by his nervousness, his date might still reasonably feel disappointed, hurt, and angry with him on the basis of his display. At the same time, it isn’t as if learning that the man’s actions did not reflect his deep
self would mean nothing. The revelation might prompt the woman to give the man another chance, or at least to moderate her lasting ill will toward him. On the other hand, it also suggests a failure of authenticity that she might take to be problematic in its own right. Of course we care a great deal about what particular actions reveal about who we are deep down; we form self-conceptions and pursue and mold relationships in large part on the basis of judgments about what our actions and the actions of others mean. But being “true to oneself” is sometimes hard, and we often fail. In such lapses of authenticity, we are still accountable for our actions, even when they genuinely do not express something “deep” about our personalities and characters.

Obviously, the cases of the waiter and the failed date will not refute all varieties of self-disclosure theories. (Theories that emphasize quality of will, for example, can account for them rather easily.) There are various philosophical moves to be made in discussion of examples like these, and I will rehearse and address them in the next two sections. My point here is that philosophers who wish to explain responsibility in terms of deep-self disclosure will need to make some moves, and that these moves will put a distance between their theorized conceptions of the deep self and the concept of the deep self we employ in everyday life.

Self-disclosure theorists have failed to fully appreciate this disconnect between the nature of our everyday interest in our deep selves and judgments of moral responsibility for individual actions. In what follows, I will argue that the responsible self is actually quite shallow, in some cases almost empty.

III: FRANKFURT AND WATSON

In the previous section, I set out to show that the ordinary sense of the deep self, though important, is not systematically connected to responsibility judgments. Expression of the morally important concept of the deep self, as it is understood in everyday life, is not necessary for responsibility and reactive blameworthiness. In this section I will canvas and evaluate some more highly theorized notions of the deep self that philosophers have tried to link directly to responsibility and blameworthiness.
The logical starting point in a survey of contemporary attempts to isolate the responsible self is the work of Harry Frankfurt and Gary Watson. While Frankfurt and Watson both tie their conceptions of the real, authentic self to moral responsibility, making this connection is not the only purpose of their work. Unlike some more recent deep-self theorists I will come to later, Frankfurt and Watson intend, it seems, to capture something close to the intuitive notion of the deep or authentic self that I articulated in the previous section.

Frankfurt is particularly interested in the concept of a person, arguing that our interest in our wills and our capacity for reflective self-evaluation are the defining features of personhood. According to Frankfurt, this caring about what we do manifests itself in the form of second-order desires, or desires about desires. One kind of second-order desire is a second-order volition. This is a desire for a first-order desire to be effective in action, a desire that we act on some first-order desire. Frankfurt holds that only persons have second-order volitions. The wanton, a man with first-order desires but no second-order volitions, does not care about his will. Genuine persons are not like this. Only persons can be free and responsible, and in free action we exercise our personhood by acting on the desires we want to be effective. To see how this works, consider the case of an unwilling drug addict. This man has a first-order desire to get high and a first-order desire to stay sober. He wants his desire to stay sober to win out, but he is overcome by his desire to get high. This person, Frankfurt explains, does not act in accordance with his second-order volition and is thus unfree.

Watson’s view is structurally similar to Frankfurt’s. He argues that we are free and responsible when we act in accordance with our values, questioning whether we need to stipulate anything about

---


105 Frankfurt, 329.

106 Ibid, 327.

107 Ibid, 330.

108 Ibid, 331.
second-order desires and volitions in order to know that certain voluntary actions are ours. What is special about our agency, Watson explains, is our capacity to value things, to judge that they are good or worth doing. When we express these judgments and act on our values, we are free and responsible. What difference does it make, he wonders, whether we have a second-order desire for some first-order desire to be effective? As long as that first-order desire is a desire for something we value, then we are free and responsible when we act on it.

These views have been discussed in great detail in the responsibility literature, and so I will try to keep my remarks brief. A few points are worth emphasizing. First, as I said at the outset, both theories are of great interest even if their verdicts on responsibility are implausible. Both accounts attempt to explore something close to the ordinary conception of the deep self, which, as we have seen, is central to judgments of identity and authenticity. Still, I am not the first to note some drawbacks of these views when it comes to their ability to explain our intuitive judgments of responsibility and blameworthiness. One classic problem involves consideration of the role of the historical origins of the self-constituting desires or attitudes. For example, if one’s second-order desires and values are the results of a neurological disorder, a trauma, or even brainwashing, Frankfurt and Watson’s theories would seem to render unintuitive, even harsh, verdicts.

Perhaps even more important is the point that rational disavowal or lack of second-order endorsement is not intuitively excusing in a large cluster of everyday cases of blameworthiness. Recall the case of Blackout Bill, the college student who gets intoxicated almost to the point of alcohol poisoning, stumbles into his friends’ apartment, and, to everyone’s shock, launches into a horrifying racist and sexist rant in which he expresses apparently deeply buried desires to perform unspeakable acts. It seems to me that Bill’s friends could accept his insistence that his words do not truly reflect his stable character. They might have no reason to think that Bill actually endorses, or in fact has ever even indulged, any of the terrible thoughts he expressed in his rant. Frankfurt’s view, which relies on the

109 Watson, 350.

110 Ibid, 350.
presence of second-order desires to ground responsible and blameworthy action, cannot make sense of Bill’s blameworthiness. After all, it isn’t clear that Bill had any second-order desires in the moments leading up to his rant, nor do I think it matters if he didn’t. Watson’s form of the view is, to my mind, similarly ill-positioned to respond to the case. It is easy to imagine that Bill’s actions were genuinely out of character, that he does not reflectively endorse the views he expressed. In other words, it is easy to imagine that Bill’s actions do not reflect his “real self,” at least in the senses that Frankfurt and Watson understand it. Though this realization might be comforting to Bill’s friends, it would still be fitting for them to feel angry, indignant, and hurt by his words, and for Bill to feel guilty and ashamed.

IV: RESPONSIBILITY AND THE SHALLOW SELF

Frankfurt and Watson’s theories have given rise to a substantial literature on the relationship between the deep self and responsibility. Descendants of Frankfurt and Watson, despite their general agreement that responsibility is a matter of self-disclosure or self-expression, diverge in their identifications of the deep, responsible self. Broadly speaking, one camp, which includes TM Scanlon\(^{111}\) and Angela Smith,\(^{112}\) argues that the true self is expressed in those actions that reflect our judgments or rational activity. Of course even among these theorists there are differences. Here, I focus on Smith’s account in particular, but I intend for my remarks to apply more widely.

On Smith’s “rational relations” theory, moral responsibility is tied to moral address. The hallmark of responsible action is that the agent can be called to account for what he has done. The intelligibility of the demand for explanation and justification, and the obligation to admit fault if these are not forthcoming, are, for Smith, the central features of responsible action. She writes:

Most of our desires, emotions, beliefs, and other attitudes seem to meet this condition of judgment-dependence, even though they do not commonly reflect a choice or decision, and are not normally under our voluntary control. These states are “judgment-dependent” in the sense that they generally reflect and are sensitive to our (sometimes hasty, mistaken, or incomplete)

---


judgments about what reasons we have, and they are generally responsive to changes in these judgments. We are ‘responsible for’ these things, therefore, because they reflect rational assessments for which we are appropriately regarded as answerable.\textsuperscript{113}

While Smith’s view, with its emphasis on rationality and judgment, is in some ways a close relative of the Watson account I discussed earlier, her conception of the responsible self stretches the notion of “judgment.” On Smith’s view, both cognitive and conative states are accountable to evaluative judgments. If properly connected to evaluative judgment actions, omissions, attitudes, and the traits that give rise to them could all be fair game for moral criticism, praise, and blame:

What matters, on this account, is whether an action or attitude is normatively connected to a person’s underlying judgments in such a way that she can, in principle, be called upon to defend it with reasons and to acknowledge fault if an adequate defense cannot be provided. Bodily movements and mental states that are not even in principle answerable to a person’s judgment are therefore not the sorts of things for which we are responsible, on this account; but we are responsible for most of our desires, emotions, beliefs, and other attitudes, despite the fact that they do not generally arise from conscious choice or decision and are not normally under our immediate voluntary control.\textsuperscript{114}

Before I respond directly to Smith’s version of the self-disclosure view, I will introduce a closely related rival. On the competing theory, the deep self is associated not with an agent’s evaluative judgments but rather with his or her conative states and dispositions. Chandra Sripada, for example, holds that the responsible self is to be identified with the agent’s cares, understood broadly to involve “a complex syndrome of motivational, commitmental, evaluative, and affective dispositions.”\textsuperscript{115} On his view, judgments about someone’s deep self, and thus about responsibility, are primarily concerned with responding to the sense the agent gives us of what matters to him.\textsuperscript{116}

In response to these dueling proposals, David Shoemaker has offered what he calls “the ecumenical view” of the real self. On this theory, the deep self is disjunctive: actions that express either

\begin{itemize}
\item \textsuperscript{113} Smith (2008), 370.
\item \textsuperscript{114} Ibid, 370.
\item \textsuperscript{116} Ibid, 1211.
\end{itemize}
an evaluative judgment or a commitment count as coming from our responsible selves and make us legitimate targets of moral appraisal.\textsuperscript{117}

I do not want to go too far into the details of any of these theories here. My criticisms do not depend on their intricacies but instead aim to strike at a more fundamental level. I argue that each of these proposals suffers from a common mistake: they assume that responsibility judgments, expression of reactive attitudes, and moral criticism are mostly about assessing a person rather than his or her action, and this is false. Determinations about the extent to which an action reflects an agent’s true personality, evaluative judgments, or cares are important to our blaming practices, but self-disclosure is not a necessary condition of blameworthiness.\textsuperscript{118} In deciding whether an agent is blameworthy for an action, we care about self-disclosure only because it affects what Scanlon has called the “meaning” of the action (its significance for affected agents),\textsuperscript{119} not because it is a necessary condition on moral responsibility. When an action does express or reveal something significant about its author, this is relevant to the content and tone of the blame we direct toward him, but the absence of self-disclosure is not fully exculpating.

Consider instances of forgetting. It is supposed to be a great advantage of self-disclosure theories that they can accommodate our intuition that we are responsible, for example, for forgetting friends’ birthdays or to pick one’s child up from school.\textsuperscript{120} In fact, self-disclosure theories fall short in their treatment of such cases, generating “blameless” verdicts for culpable agents. For now, consideration of

\textsuperscript{117} Shoemaker, D. (2015). Ecumenical attributability. In The Nature of Moral Responsibility: New Essays, ed. R. Clarke, M. McKenna and A. Smith, 115-140. New York: Oxford University Press. Shoemaker is concerned specifically with aretaic appraisal, but this will not be important to my criticisms, which will show that some actions that express neither an agent’s commitments nor his evaluative judgments can be subject not only to aretaic appraisal but even to strong forms of reactive praise and blame.

\textsuperscript{118} My arguments are meant to apply to the entire spectrum of self-disclosure views, including, for example, theories like the one offered by Nomy Arpaly, which takes the expression of deficient quality of will to be necessary for blameworthiness. For a defense of this view, see Arpaly, N. (2006). Merit, Meaning, and Human Bondage: An Essay on Free Will. Princeton, N.J.: Princeton University Press.


\textsuperscript{120} Smith, in particular, makes significant use of these examples.
Shoemaker’s proposal will be most instructive. I will show that even the most “ecumenical” theory of attributability as self-disclosure has the potential to distort our reactive judgments.

Recall that for even the most inclusive self-disclosure theorists, what matters to an agent is central, and that to be blameworthy for an action the action must be evidence that one’s commitments, cares, or judgments (ways of mattering) are somehow deficient. The self-disclosure strategy in the forgetfulness cases, then, involves the claim that the parents and friends in the examples are blameworthy because their evaluative judgments about the importance of the birthday or their level of caring about their children open them up to moral criticism.

Surely this is true of some forgetting cases, but need it be true of all of them? Is every blameworthy instance of forgetfulness traceable to some genuinely “deep” shortcoming? I think not. We can easily imagine cases in which one just forgets despite genuinely caring deeply, though not obsessively, about a friend’s birthday. Or consider the following case: Johnny has agreed to pick up his friend Camila at the bus stop. In the morning, she calls to let him know that she decided to catch an earlier bus because of concerns about a snowstorm that is expected to hit in the afternoon. Right after he hangs up the phone, Johnny gets word of a relative’s medical emergency. He spends the day talking to various family members, is totally preoccupied, and, because the snowstorm never hits, is never reminded of the early pickup. A few hours later, he gets a call from Camila and realizes, to his horror, that he has left her stranded outside in the cold!

Is Johnny blameworthy for his forgetting? A self-disclosure theorist might argue that he is, because his forgetting reveals an objectionable lack of concern for his friend. But does his forgetfulness necessarily show anything of significance about his cares, evaluative judgments, or quality of will? Perhaps, in some version of the case, but I don’t think we need to answer “yes” in order to conclude that it would be fitting for Camila to feel, and even express, some level of anger, upon his arrival, and even after she learns the full explanation of his tardiness. After all, Johnny has no excuse; no one made him forget to make the pick-up—the mistake was entirely his fault. As a result, he owes Camila an apology despite
the fact that the afternoon’s events revealed, if anything, only that he has a very specific disposition to forget this particular thing under these highly unusual conditions.

At the most general level, self-disclosure theories are liable to the same kind of criticism I leveled against control theories in the previous chapter. They fail because they posit a strict metaphysical condition on responsible and blameworthy action that does not reflect the actual norms of justified blaming practices; thus, when the condition is not met, the theory automatically generates blameless verdicts for blameworthy agents. My sense is that it is fitting for Camila to be upset upon Johnny’s eventual arrival because blame and indignation are often responsive to agents’ failures to fulfill their obligations or meet reasonable expectations. Johnny has let his friend down. He has failed in a duty; to go on without an apology or an expression of remorse would be to devalue Camila, to send a message that he is socially superior, above reproach. And it is not merely that it would be nice of Johnny to apologize; it is, rather, that an apology is owed: even though his omission doesn’t necessarily reveal anything “deep” or interesting about him (his judgments, his cares, the quality of his will, etc.), he has wronged his friend. He made a mistake and he’s accountable for it, just as he would have deserved praise, or at least a “thank you,” had he executed his duty as promised.

Consideration of guilt and apology in these contexts can help to bring out the point. When we let down people we care about because of a failure or mistake that is entirely ours, feelings of guilt can reveal an investment in the relationship, a fitting discomfort in knowing that our error has harmed another, and a desire to do better in the future. The expression of these feelings can be significant for a relationship after this kind of non-disclosive failure or mistake; in the absence of an apology, the wronged party may feel that he is diminished, that his pain has been forgotten or overlooked, and may be justified in employing angry blame as a means of eliciting a gesture of remorse.

To be clear, this does not mean that the lack of self-disclosure is irrelevant in such cases. Quite the opposite: the fact that Johnny does care a great deal that his omission led to his friend’s being marooned at the bus stop in the cold is something he would no doubt emphasize when he finally arrived. All I am objecting to is the idea that the non-revelatory nature of the forgetting in this case gets Johnny
off the hook completely. It does not. He really *does* care about stranding Camila, and this is precisely what makes him feel (appropriately) remorseful and guilty, two hallmarks of blameworthy behavior. Often, angry feelings in these kinds of cases will be short-lived (especially if an apology is offered), although not always. After all, the consequences (and a host of other factors) also affect the meaning of the action and the kind of reactive response that is warranted: if Camila gets frostbite waiting for Johnny to pick her up, then her justified blaming reaction may be harsher because of the event’s greater seriousness and significance.

But, a self-disclosure theorist may object, even if one grants that Johnny’s forgetting does not reflect a judgment that Camila isn’t important, surely it *does* reflect a judgment about the significance of his relative’s medical emergency. 121 And because the forgetting is traceable to *this* judgment, and the resulting preoccupation, the self-disclosure theorist may argue that her theory can account for Johnny’s blameworthiness after all.122

It is worth noting that this response trades on a subtle but substantial shift on the part of the self-disclosure theorist. So far, we have been working under the assumption that on the self-disclosure model, blameworthiness requires the expression of an *objectionable* or *deficient* evaluative judgment.123 In the case at issue, however, all of Johnny’s judgments (assuming, as we have been, that he cares both about his relative’s situation and about Camila’s welfare) are appropriate, and even commendable. There is no deficient evaluative stance, attitude, or judgment upon which to ground his blameworthiness.

But even a more relaxed version of the self-disclosure theory, on which blameworthy action may be traceable to *admirable* evaluative judgments, will struggle to explain a range of forgetting cases.

---

121 For simplicity, I will focus on judgments in discussing this objection, but the response applies to care-based views as well.

122 I am grateful to an anonymous reviewer from *Philosophical Studies* for raising this objection.

123 Smith (2005) seems to suggest this version of the view. She writes, for example, of someone whose contempt for members of a racial group stems from a judgment about their intellectual inferiority that “the person is open to rational, and in this case moral, criticism for this attitude precisely because of its rational dependence on these objectionable underlying evaluative judgments” (254).
Imagine, for example, a slightly different iteration of the Johnny and Camila example: After Camila calls to change the pickup time, Johnny receives an email from a coworker linking to a mildly amusing YouTube clip rather than a phone call about a sick relative. One video leads to another and, eventually, Johnny loses track of the time. From here, everything else is the same: the snow storm never hits, he isn’t reminded of the early pickup, he strands Camila.

Is the forgetting in this modified example traceable to an evaluative judgment of Johnny’s? My contention is that it need not be for him to be responsible and blameworthy for his lapse. For Smith, “[Evaluative] judgments, taken together, make up the basic evaluative framework through which we view the world. They comprise the things we care about or regard as important or significant.”124 But to claim that anything of even moderate significance is necessarily disclosed in this case seems overly hasty. In the modified example, what disturbs Johnny’s routine is a sequence of random, banal distractions, combined with the change to the pickup schedule he had originally planned for. His forgetting need not be traceable to any evaluative judgment at all. Left without a judgment of even fleeting significance to stand on, the self-disclosure theorist, should she wish to take this route, may be forced to conclude that Johnny’s blameworthiness for his forgetting in the YouTube version of the case is somehow more open to question than it is in the original story, in which he is distracted because of news that he rightly cares deeply about. This would be an odd result.

At this point, one may be tempted to respond that Johnny’s forgetting, especially in this newest iteration of the case, must show something about him after all. Could someone who truly cared about his friend have zoned out watching internet videos? In short, yes. I contend that these sorts of non-revelatory lapses are indeed possible (and even relatively common!), and that we may be rightly blamed for them when they occur.

At the outset, I suggested that self-disclosure views can distort our understanding of blamed agents. I am now in a position to explain myself more fully. As we saw in section II, we are deeply invested in the everyday notion of the self. The idea is central to our forming self-conceptions, to

124 Smith (2005), 251-2.
understanding our relationships with others, and to shaping the narrative structures of our lives. Those who hold self-disclosure views of responsibility risk warping these processes. I have suggested that the forgetful person might be blameworthy without revealing anything significant about his evaluative judgments or his cares. I have stressed the oddness of giving up on blame in cases like these, but saving the reactive attitudes by insisting that these lapses are self-disclosive after all may be even worse. We care a great deal about what actions tell us and don’t tell us about actors, and to insist that all blameworthy mistakes, reactions, lapses, and failures of memory reveal something deep about us, our judgments, or the quality of our wills could lead to potentially significant distortion; the more invested in the importance of the self we are, the warier of the inference from blameworthiness to “deep” deficiency we ought to be. To claim that Johnny’s blameworthy lapse then, need be revelatory of his judgments, cares, or quality of will, is to invite a mistaken judgment about his character (or, at least, his judgments/cares). He is responsible and blameworthy for his forgetting because he stranded his friend and has no good excuse; what the incident actually reveals of his judgments, cares, and character is an important further question that deserves to be investigated and answered on its own terms.  

These cases are closely related to another group of counterexamples to self-disclosure views of this kind: mistakes. It may be helpful to think first about a recent non-moral example: When Minnesota Vikings kicker Blair Walsh missed a short field goal attempt that would have sent the Vikings to the second round of the 2015 NFL playoffs, he said: “It’s my fault…. I’m the only one who didn’t do my job there. So that’s on me…. I worked real hard to get myself to a place where I was very consistent for this team all year, and in that moment, the moment they needed me the most this year I wasn’t and that stings…. I’ll take the blame because I deserve every second of it.” Is Walsh guilty of any failure of  

---

125 Or our judgments, cares, etc.

126 This is not to say, however, that the two questions are entirely independent. As I have said, the sort of blaming reaction that will be fitting in response to Johnny’s forgetting will vary depending on how revelatory of bad character the incident is. What I am objecting to is the claim that disclosure of anything significant is required to cross the blameworthiness threshold.

caring, evaluative judgment, or quality of will here? It seems unlikely. After all, he seems to care deeply about the team and to have had an excellent work ethic that reflects his exemplary level of investment. One could also claim, I suppose, that his miss was a failure of football-related evaluative judgment. While this may be true of some version of the case, we can easily imagine that no evaluative judgment was involved at all, that Walsh simply acted. *He just missed.* Still, despite his purity of evaluation, commitment, and will, Walsh’s teammates and coaches might reasonably be upset with him: he should have made the easy kick, and his misfire cost his team the game. As he put it in his postgame remarks, he didn’t do his job, and, as a result, he must “take the blame.”

We can construct similar, graver, cases on the same model. A doctor tries to do her very best but makes a simple mistake during surgery. An accountant, even after a thorough double check, makes an oversight that costs a coworker her job. These agents, given their roles, are obligated to do their jobs correctly. The fact that their errors do not reflect misguided values, lack of caring, or substandard quality of will means that their actions aren’t as bad as they could be; nevertheless, it is fitting when people they harm feel and express some degree of anger toward them.

For the self-disclosure theorist, as we have seen, these cases pose a dilemma: either the forgetting, erring, or failing person betrays a deficiency of caring, judgment, or quality of will, or else she is not blameworthy for her behavior and its results. I have rejected this dilemma, but I should say a bit more about the theoretical underpinnings that give it its intuitive appeal.

Philosophers of moral responsibility have found themselves in this position, I think, because of an antecedent attachment to a theoretical picture on which blame and responsibility depend entirely on

---

128 In response to this case, a Smith-style self-disclosure theorist may note that the accountant in question *does* make a faulty judgment of a certain kind: she should not have judged, say, that the books were balanced when in fact they were not, or that the paperwork was suitable for submission. While this is one way of describing the situation, it is not in the spirit of Smith’s use of the term “evaluative judgment.” Smith is focused on those judgments that one can, in principle, be called upon to justify and that involve one’s “‘taking’ or ‘construing’ things to be a certain way or to have a certain significance” (Smith (2005), 260). To interpret clerical errors as involving such judgments seems a stretch; it would be odd to ask someone to justify his overlooking a number on a spreadsheet, or his mistakenly writing that the sum of two and two is five. It *does* make sense to ask one to justify his decision not to double-check his work, but I have been pointing out that mistakes can occur even when one *does* double-check, and arguing that we are often still responsible for them when they do. Thank you to an anonymous reviewer from *Philosophical Studies* for pressing me to clarify this point.
metaphysical or psychological features “internal” to the blamed agent, and which suggests a theory of blameworthiness on which an impartial observer doles out demerits in perfect proportion to the “stains” on the agent’s soul.\textsuperscript{129} While many contemporary self-disclosure theorists would no doubt denounce such a conception of blame and blameworthiness, one can see the remnants of this picture in almost every prominent theory on offer in the contemporary literature. The relic of this understanding of blameworthiness is the self-disclosure condition itself, the presence of which ensures that blameworthiness depends upon, and that blame responds to, some failure of the blamed agent’s soul, self, character, judgment, or quality of will.

Of course, in many cases it seems obvious that blame is responsive to the bad features of a wrongdoer’s character that are manifested in his or her action. But even in cases in which the expression of ill will is essential to understanding blame, the blaming reaction is shaped and textured by other factors that affect the action’s meaning: consequences, relationships, and context.

The implausibility of theories that demand that the harshness of our reactions to bad behavior should perfectly track a blamed agent’s level of controlled wrongdoing, or poor quality of will expressed in his or her action, has been apparent to laypersons and philosophers alike. As we saw in chapter three, it is immediately obvious that even the most innocuous uses of “tracing” procedures by which one “traces” an agent’s blameworthiness for an action or outcome back to some earlier failing seem to license the expression of harsher reactive attitudes toward unlucky agents than toward their lucky counterparts. In one sense, we might say that two drunk drivers, both of whom demonstrated the same lack of good will and failure of caring and judgment when they got into their cars after a night of drinking, for example, are “equally blameworthy,” despite the fact that one driver was unlucky and injured a pedestrian and the other returned home without incident. But when it comes to the responses sanctioned by our moral practices, the character and tone of the reactive blame that is appropriately directed toward the drivers is subject to huge variation depending on what happened on their trips home.

If one is willing to concede that reactive response ought to vary at all in this way, I think that one ought also be willing to entertain the thought that requiring even a minor deficiency of will or character in order to get justified blaming reactions off the ground may also be a mistake. In other words, one should be ready to take seriously the contention that the sorts of faults or mistakes that the reactive attitudes address (and ought to address) are not confined to deficient psychological states. One lesson of moral luck is, of course, that our justified responses are shaped in part by the outcomes of blamed agents’ behavior. But this insight suggests another lesson that may be even deeper and more important: namely that we are not always interested in reacting to psychological states or motions of the will at all; the reactive attitudes respond to actions at a different level of description.

We react to actions as events in an ecosystem, not as internal phenomena. Interpersonal expressions of resentment and blame are social responses; when we blame, we describe the blameworthy behavior in “thick,” inescapably social, terms. One is blamed for hurting another’s feelings, for stranding a friend at the bus stop, for botching a surgery, for getting a colleague fired. These are the socially disruptive “actions” that reactive blame is equipped to address.  

This is not to say that blameworthiness does not require some wrongdoing or lapse on the part of the blamed agent. It does. But once we describe wrongdoing in the “thick” social terms I have suggested, we can begin to take a view of blameworthy behavior that is broader than the narrow, “internalized,” conception of blameworthy failure imposed upon us by self-disclosure views. This understanding of the ways in which justified reactive practices work can help to explain, not only the persistent judgment that contingent factors like consequences should affect blame, but also why actions (as opposed to mental states or the internal workings of the will) are the loci of blame in the first place.

I pursue this suggestion in greater detail in chapter five.

For this reason, the point is distinct from Bernard Williams’s discussion of the “lorry driver who, through no fault of his, runs over a child,” and feels agent-regret, or Adam Smith’s consideration of “a man of humanity, who accidentally, and without the smallest degree of blamable negligence, has been the cause of the death of another man, [and] feels himself piacular, though not guilty.” In the cases I am interested in, there is fault, and so guilt and blame, rather than agent-regret, are fitting. See: Williams, B. (1981). Moral Luck. Cambridge: Cambridge University Press, 28, and Smith, Adam. (1976). The Theory of Moral Sentiments. Oxford: Clarendon Press, 107.
We are now in a better position to evaluate Smith’s claim that “Moral criticism, by its very nature, seems to address a demand to its target. It calls upon the agent to explain or justify her rational activity in some area, and to acknowledge fault if such a justification cannot be provided.”¹³² Consideration of the foregoing cases suggests that Smith’s thesis is true of some forms of moral criticism but does not tell the whole story. Moral criticism, and our blaming practices more generally, are not so straightforward. When we engage in these activities, we respond not only to personality defects, failures of caring or judgment, and lapses in quality of will; we engage in a process that allows us to express a wide range of feelings that are responsive not just to individual elements of the target’s psyche or their expression in action, but to the whole state of affairs those actions gave rise to. As we saw in chapter two, these responses allow us to stand up for others and ourselves, restore social order, articulate standards of conduct, and, if things go well, to start an interaction that could prompt the target to feel the self-inflicted pain of guilt and offer an apology. The blaming process does not end with, nor does it aim at, a mere acknowledgement of fault, at least in the narrow sense of “fault” that the self-disclosure theorists I have discussed seem to have in mind.

Before concluding, I wish to address one more variant of the deep-self approach suggested by Timothy Schroeder and Nomy Arpaly.¹³³ Schroeder and Arpaly offer what they call a “whole-self” view in response to hierarchical deep-self views that privilege one aspect of the self (rational judgment, for example) over another (say, desire). Many of the arguments I have offered here apply to the whole-self view, as it is in some ways quite similar to Shoemaker’s “ecumenical” proposal. But in their discussion of the whole self, Schroeder and Arpaly make some interesting remarks on the integration of the self that merit additional discussion.

¹³² Smith (2008), 381.

First, this claim: “Other things being equal, an agent is more praiseworthy for a good action, or more blameworthy for a bad action, the more the morally relevant psychological factors underlying it are integrated within her overall personality.” In many cases, this seems plausible. Again, consider Blackout Bill: Bill does not reflectively endorse racist and sexist thoughts, and I argued that his high level of blameworthiness is, in large part, due to bad luck. But what if Bill frequently indulged the fantasies he expressed in his diatribe and secretly resented women and racial minorities? It seems to me that this second version of Bill might merit harsher blaming reactions than his unlucky counterpart, and it is plausible to explain the difference in fitting reactive response by appealing to the way the rant fits with, and gives voice to, the racist and sexist elements of Bill’s stable character.

In other cases, though, the integration principle seems to fail. Imagine two men who conspire to defraud patrons of local restaurants by stealing their credit card numbers. Neither man needs the money, and both have committed this sort of crime countless times before. In fact, the only difference between the two men is to be found in the thoughts that run through their heads as they commit their crimes. The first man experiences no internal conflict and always executes the job smoothly. The second man is always struck by the hardship he is about to inflict upon the local restaurant owners and their employees every time he starts to steal the credit card numbers. Then, the thought passes and he executes the job smoothly. Who is more blameworthy for the fraud? The answer is not at all obvious. To my ear, the most plausible reading of the case is that the first man is the worse man, but that the second man’s actions may be deserving of a harsher reactive response. After all, the fragmentation of his psyche forced him to confront the suffering he was about to cause…and he did it anyway! At the very least, the matter should be up for discussion. It is far from obvious that the more divided man deserves less blame than his wholehearted friend. In fact, there is no reason to assume even that the two merit the same kind of reaction.

Certain cases of praiseworthy actions cause a similar problem. Consider the case of Huckleberry Finn’s decision not to turn Jim in that is central to Schroeder and Arpaly’s argument against hierarchical

134 Ibid, 172.
views of the self. They write: “Of course, Huckleberry would be more praiseworthy still if he had no racist beliefs at all – if he were more integrally good – but he may merit high praise nonetheless.”

Once again, this doesn’t strike me as obvious: Perhaps integrated Huck is a better person overall, but the internal conflict of Twain’s original makes his action especially praiseworthy. In order to do the right thing, he had to choose friendship and compassion over an incorrect, but settled, moral judgment…, and this is impressive! A more integrated Huck’s parallel action would be worthy of praise as well, but perhaps not to the same degree, and perhaps not even of the same kind.

Second, and more directly related to a central theme of this paper, is the Schroeder/Arpaly tenet of the necessity of self-expression for responsible and blameworthy action. Schroeder and Arpaly write:

It is open to a Whole Self theorist to allow poor integration below some threshold as an excusing condition on moral responsibility […]. An act which is extremely poorly-integrated may be one that, as one might say, is better attributed to circumstances, a drug, or a nefarious neurosurgeon than to the actor, and so be an act for which the actor is not responsible.

One way of putting the main contention of this paper is that this sort of integration is not a necessary condition of responsibility and blameworthiness. I have argued that responsible action can, at times, arise from an extremely “shallow” part of the self: in some cases the only thing expressed by responsible action is the trivial fact that the agent was disposed to act in a certain highly specific way under certain highly specific conditions.

The foregoing points steer us toward a richer understanding of what guides justified reactive feedback. Such an understanding, which can only be reached via consideration of our justified ethical judgments and reactive practices, will leave us with a conception of responsible and potentially blameworthy agency that is much wider than philosophers have imagined, and will save the important everyday concept of the deep self from the encroachment, and potential distortion, of our theories of moral responsibility.

135 Ibid, 178.

Still, one may feel as though we are left with the problem we began with: On what grounds can we draw the boundaries of the responsible self? If anything, the problem now seems even more vexing: if the responsible self can be so shallow, and its activity so unrevealing, what hope do we have of systematically ruling out spasms, neuroscientist interventions, and the like?
I: INTRODUCTION

The previous chapter ended on a low note. I showed that self-disclosure theories, despite their ability to account for the blameworthiness of agents like Blackout Bill, whose uncontrolled behavior violates social rules and disrespects others, cannot explain examples in which agents make blameworthy mistakes without expressing anything significant about their “selves,” quality of will, judgments, or cares. This conclusion left us staring at an even darker vision of the problem we set out to address: with a conception of the responsible self that is so shallow, and whose activity is so unrevealing, on what grounds can we mark out responsible, and potentially blameworthy agency? Why is Johnny’s forgetting to pick up Camila at the bus stop any more blameworthy than a sudden, uncontrollable muscle spasm? Why is anger an apt response to the missed field goal and the surgical error, but not to Bernard Williams’s unlucky lorry driver who, despite following the rules of the road, runs over a child?137

In this chapter I endeavor to answer these questions. To do so, I propose a reorientation of focus, away from questions of attributability or action ownership, and toward the social practices that govern the expression of the reactive attitudes. I argue that the question of attributability, of whether an action is related to an agent in a way that renders him eligible for reactive feedback, plays a far less significant role in determining whether others may blame him than philosophers of moral responsibility have previously imagined. I show how excessive focus on action ownership leads to the mistake I highlighted in the previous chapter: the assumption that blame is about the agent rather than his or her act.

Finally, I defend a rule-based understanding of the norms that govern ethical response, and present a positive proposal for distinguishing between blameless and blameworthy instances of non-expressive agency. I conclude that the responsible self is wide and shallow, and that the attributability

necessary for responsibility need not go beyond the commonsense notion captured by expressions such as “It was me.”

II: TWO BARRIERS TO BLAMEWORTHINESS

Very roughly, we might understand the justified expression of a reactive attitude such as praise or blame toward an agent as requiring at least the following two things. The first is the metaphysical or conceptual requirement that the action in question “belong to” or come from the agent being addressed. It is unfitting to blame me for the actions of someone I have nothing to do with. The second is the explicitly ethical requirement that the reaction be an apt response to the action. It is unfitting and wrong, for example, to blame someone for an excellent deed. That the first of these conditions is met is plainly not enough to settle what sort of response is warranted. We are the “authors” of all sorts of actions; ethics is required to determine how to react to any one of them. That the second condition is met is not enough to justify a reaction on its own, either; we must direct the feedback toward the right person. It is wrong and nonsensical to blame someone else for my wrongdoing. How, exactly, is the philosophical idea of moral responsibility related to these two conditions?

It seems natural to say that such a theory would underlie the first requirement, that the target of a reactive attitude on the basis of some action must be connected to that action in some significant way. If this is the proper place for a theory of responsibility, then to claim that someone is responsible for an action is to say only that the action may be properly imputed to him, that he “owns it.” It is not to say that any reactive attitudes may be appropriately felt or directed toward him simply on the basis of the act

---

138 Or, for the grammar-conscious, “It was I.”

139 At least usually. A parent may take the blame for a child’s action, for example. I will not discuss cases of that kind here.

140 From now on, for simplicity, I will refer only to actions, though what I will say will apply to other behaviors, attitudes, and omissions one may be responsible for as well.
being “his,” although responsibility is a necessary condition for legitimate feeling and expression of reactive attitudes.

While reactive attitudes are often appropriately directed toward responsible agents, reactive aptness is not settled by responsibility: an agent can be responsible for a bad action (the action is attributable to him) without being a fitting object of negative reactive attitudes in response. Once the attributability barrier has been crossed and the agent is deemed responsible, we can begin to decide how to react to him on the basis of ethics of blame, which will be informed by what T.M. Scanlon has called the “meaning” of his action, or “the significance of [the] action for the agent and others.”141 Of course, the nature of the act itself and the psychology of the actor will affect its meaning, but other factors, such as the context in which the act occurred and the blamer’s relationship to the blamed agent, will be relevant as well. Settling the question of proper reaction to an action will happen, as Angela Smith writes, “at the level of moral response, not at the level of basic moral responsibility.”142

This structure of a theory of responsibility and blameworthiness strikes me as intuitive, but it isn’t always clear whether the most prominent theories on offer fit this shape. Consider the control theory, whose key principle is that we may only be legitimately blamed for what we directly or indirectly control. Which aspect of blameworthiness—attributability or the ethics of blame—is a lack of control supposed to undermine? The answer is not obvious. Perhaps the most natural response available to the control theorist is that it undermines attributability: the lack of control severs the connection between agent and action necessary to make him a proper target of reactive attitudes. But, as I argued in chapter three, leaving it at that is unsatisfying; we want to know why the lack of control blocks action ownership.

In chapter three I also argued that if the control condition is to have any plausibility, it must be understood as an ethical principle. When agents like Blackout Bill, for example, express deeply-held attitudes in a way that disrespects and diminishes others, it is not a sign of confusion or conceptual error to express anger and resentment in response. Implicit in this thought is the idea that Bill’s ranting is


attributable to him.\textsuperscript{143} Thus, to survive, the control principle must be given an explicitly ethical reading; it must assert that blame for the uncontrolled is morally wrong or unfair. In other words, a lack of control would have to undermine the ethics of blame directly, and not by way of concerns over attributability.

What about the self-disclosure approach? The basic blame-related principle of the self-disclosure theory is that we may only be blamed for those actions that express our judgments, cares, or quality of will. Which blameworthiness barrier blocks non-disclosive action? Here, a more plausible case can be made that non-expressive action is meant to fail the attributability test. Attributability is a question of the connection between an agent and his action, and non-disclosive actions (obviously) do not express anything of significance about their authors. But who is to say that an action-agent relationship characterized by significant expression must hold if blame is to be justified? Again, the question cannot be answered without some substantive understanding of the norms, purpose, and \textit{modus operandi} of the reactive attitude in question (in this case I will focus on blame, but the analysis will vary depending on the reaction). In order to answer the question of what can be attributed to an agent, we must first know why we are asking. This means that the interesting questions about whether someone is blameworthy in a particular case will arise in the realm of the ethics of moral response rather than in the realm of attributability. The ethics of blame are primary: without an understanding of what we are doing when we react, we cannot begin to set attributability standards.

This fact is somewhat obscured by self-disclosure views such as Smith’s, which run the standards of attributably together with a theory of blame’s purpose. For Smith, attributability requires self-expression because she understands blame as “moral criticism” and “a direct challenge to [someone] as a moral agent.”\textsuperscript{144} Attributability is analyzed as answerability, and from there the conclusion that we can

\textsuperscript{143} As we saw in chapter three, Bill may wake up the next morning and insist that “that wasn’t really me,” but it isn’t conceptually confused disagree with him. After all, the thoughts expressed reflected his deeply held feelings.

\textsuperscript{144} “Moral criticism in general, I would argue, can only be directed to a person with regard to things that involve her rational activity in some way. Our physical traits and natural abilities, in themselves, do not seem to reflect our rational agency or activity in a way that would make them an appropriate basis for moral criticism.” (Smith, Angela. (2008). Control, Responsibility, and Moral Assessment. \textit{Philosophical Studies} 138(3), 367-392, 381).
only demand someone justify or answer for those things that are governed by their evaluative judgments and rational activity, is easy. She writes:

If I am called ugly, I may feel embarrassed, amused, or incredulous, but what I am not likely to feel is challenged: that is, I will not see any reason to try to defend myself against the charge (or to apologize or make excuses for my alleged unsightliness). If I am called selfish, by contrast, I cannot regard that as merely an unwelcome description of some trait or feature I possess: it is a direct challenge to me as a moral agent, and therefore implicitly calls for some kind of response on my part. This is not to say, of course, that I will necessarily acknowledge or take seriously such a challenge; it is a point, rather, about the nature of moral appraisal itself, and how it differs from mere negative description.145

The passage gives us an insight into how one’s understanding of the purposes and functioning of blame can shape one’s theory of attributability. Smith contrasts critique of one’s physical appearance with accusation of bad character, characterizing the former as “unwelcome description” and the latter as “direct challenge” to a person as a moral agent. These are different kinds of criticism, to be sure, but to suggest that they are the only, or even most relevant, forms of critique would be misleading. Both “unwelcome description” of the kind Smith has in mind, and “direct challenge” to a person as a rational agent, that is, challenge to her evaluative judgments and rational activity, attempt to get at something relatively stable about their targets. But there is another option, one that I think best captures the kind of criticism blaming responses characteristically express: the demand for apology as a response to rule-breaking. This reaction to bad behavior is neither an “unwelcome description” nor necessarily a “direct challenge” to the agent as a reasoner. If this understanding of blame is correct, it will give us reason to question the dominant assumptions about what a standard of attributability must look like.

III: BLAME AS A LEVELING RESPONSE TO WRONGDOING

In chapter two, I argued that the characteristic function of blame is to help correct imbalances in relative social rank that arose because of the blamed agent’s behavior and its consequences. I suggested that when expression of angry attitudes is successful, it can, in one sort of ideal, “rationalized,” case, prompt the target to reconsider his or her behavior and to feel the pain of guilt and remorse. These

145 Ibid, 381.
feelings, in turn, may then prompt the offender to humble himself via the ritual of apology, reaffirming his commitment to the relationship and its standards of conduct, and professing a desire to mend or improve the relationship and proceed on grounds of moral equality. In other cases, blaming can be successful if the transgressor backs off or changes his behavior, or simply because the blamer is able to express, with dignity and authenticity, her willingness to fight for respect, to maintain or restore a relationship, or to support someone else.

Imbalances in relative moral status arise when one person in a relationship or a social group gets away with violating justified rules that structure interactions between parties. When one person is “above the law,” there is a de facto disparity in standing that the others may reasonably resent. Blame is a way of responding to wrongdoing and its results, to damage done because of rules violations or flouted expectations. It also allows wronged parties, or others who have a stake in the relationship or system whose rules have been violated, to express their concern and pain in an authentic and forceful way. Blame works, I argued, because it is often scary and jarring: it is a visceral promise to take action if the wrongdoer does not humble himself and assure the wronged parties of his investment in the relationship and his willingness to be governed by its standards.

Already, one can see differences emerging between this understanding of blame and Smith’s, which contrasted unwelcome criticism (calling someone ugly, for example), with challenges to a person as a rational agent. It seems to me that neither of those descriptions adequately captures the central role of blame as a response to rule violation. A great deal of angry blaming does, of course, challenge the wrongdoer as a rational agent, but not all blaming fits that mold. In chapter four, I made the case that some blameworthy behavior is entirely non-expressive; it tells us nothing of significance about its author. In those instances, we are, I suggested, responding to the act, and not the agent. Smith’s conception of the functioning and purpose of blame, as a means of direct challenge to a person as a reasoner that demands a rational justification and response, understandably pushes us away from this possibility. This is a deep difference: If blame is always about criticism of the agent himself, we cannot make room for the thought that sometimes blaming is hardly about the agent’s reasoning at all.
But neither is blame mere “unwanted criticism.” This alternative, as Smith describes it at least, is still too personal. According to the conception of blame that I have been defending, blame is not fundamentally about critiquing the wrongdoer, but rather about restoring equality of status to the wronged. We don’t apologize for who we are; we apologize for what we’ve done. Here, I follow Herbert Morris’s remarks on the differences between blame and shame. The morality of blame and guilt is what Morris calls a “threshold” morality. That is, it is concerned with violations of a minimum standard of conduct. As a result, blame requires fault or wrongness, the failure to “meet the demands that others might reasonably place on one.”  

Shame, on the other hand, is “scalar”; it responds to a failure to realize an ideal of excellence. As Morris puts it, “What is valued in a shame morality is an identity of a certain kind and not, as is necessary with guilt, a relationship with others.” Blame is not a negative description, nor even necessarily a criticism (though blaming responses can be personally critical, which is why being blamed can often make wrongdoers feel ashamed). This is because blame is an assertion of wrongdoing and a demand for a leveling gesture (usually an apology) that conveys a commitment to a relationship and the rules that govern it. This explains the persistent intuition I appealed to in chapter four—namely, that it sometimes makes sense to blame people while knowing full well that their actions do not reflect anything significant about them.

IV: RULES, EXPECTATIONS, EXCUSES

The conception of our blaming practices that I endorse relies heavily on the idea of fault, understood as the violation of relationship-governing rules of conduct. The reason Johnny is blameworthy for stranding Camila, the doctor for his surgical slip, the kicker Blair Walsh for his shocking miss, and the accountant for her costly clerical error, is because they have violated reasonable expectations for people in their positions. They have fallen short of the relevant station-specific minimum standards we have established.

---


147 Ibid, 61.
But minimum standards are complicated by the possibility of excuse. If Johnny gets a flat tire on his way to the bus stop, or if a bird flies into the ball’s path after Walsh boots a picture-perfect kick, or if the accountant’s mind is suddenly taken over by a mad scientist, these agents would all fail to execute their tasks. And yet everyone agrees that these circumstances excuse them from blame. Why should that be the case when non-disclosive forgetting, missing, and overlooking, do not? The self-disclosure theorist has a seemingly plausible explanation: there is nothing Johnny, for example, could have done about the flat tire, and so his failure to make the pick-up says nothing about him. To be clear, control and self-disclosure proposals at this level must not be about attribution—to be plausible, they need to be understood as ethical claims about the value of a certain kind of rule that governs what counts as an excuse. This sort of self-disclosure claim, for example, would amount to the contention that our excuse-rules ought to be sensitive to self-expression at a granular level: you are excused in every instance in which your failure to meet the relevant minimum standard expresses nothing of significance about you, your judgments, or your cares.

Of course, such a rule would be revisionary in a wide range of cases. In chapter four, I noted that we do not normally take non-disclosive failures of memory, lapses, errors, and mistakes to be excusing. But is this a good excuse-rule? Would the more precise alternative make for a better blaming practice? The question is difficult to evaluate. Arriving at a clear answer would require a comparative evaluation of the different systems along various practical dimensions. I cannot provide a full accounting of that kind here, but I will make some points in favor of an idealized system of excuses roughly in line with the one we have.

In order to do so, I should first say something about why failures of memory, mistakes, and lapses do not, and should not, normally excuse, and to do that I must make a few comments about rules. Rules that are meant to set minimum standards of conduct, are, in general, supposed to guide our actions. It is not wrong, as we saw in chapter two, to blame people for what they cannot control, but it is impractical and ineffective to make a social rule that people cannot obey. Rules worth having are, in general, ones we can will ourselves to meet the reasonable expectations it sets out. A person who violates a first-order rule
with the kind of valid excuse I am interested in in this chapter does not failed to live up to reasonable expectations. Excuses are complications of first-order rules that help to set such expectations; they are meant to cover classes of cases in which would-be transgressors could never be expected to conform their wills to follow the rule.

This outline of rules and excuses is obviously far from complete, but it is enough to give us a start toward understanding why, say, getting a flat tire on the way to the bus stop is exculpatory for Johnny when his forgetting is not. The difference stems from the fact that forgetting is the sort of thing that, in many cases, does implicate one’s will. We can employ all sorts of techniques to help us remember to do things, and we tend to remember the things we really care about. A flat tire, on the other hand, and similar freak intrusions of outside forces, are not like this. A bird suddenly flying into the path of the football never reflects poorly on the field goal kicker; no one could ever expect a museum patron to will himself to avoid an unpredictable spasm in the ancient ceramics exhibit.

It is a common intuition that the impossibility of willing oneself to meet the usual standards of conduct makes blaming these agents unfair. In fact, it would be unfair, but it is worth taking a moment to revisit an argument from chapter three in order clarify the nature of the unfairness at issue. The unfairness of blame for these agents does not stem from some metaphysical principle that links control and blameworthiness. Rather, blame for these agents is unfair because a justified excuse-rule gets them off the hook. Blaming them would violate this rule, and would therefore be unfair in the same way that it would be unfair to call a runner “out” when he reached base prior to being tagged in a game of baseball.

Perhaps at this point one might be tempted to say that the rules that excuse the spasm and Flat Tire Johnny’s tardiness are good because rules that did sanction blame in these cases would be unfair in the more abstract, metaphysical sense. I think we should resist this temptation. It is not metaphysical unfairness that would delegitimize rules of this kind, but rather the simple fact that they could never do their jobs as effective rules. They could never guide action. The circumstances are so unpredictable and unavoidable that there would be no point in deterring, or encouraging people to avoid, the problematic behavior being blamed.
Forgetting, oversight, and error, however, are different. We can, and often do, go to great lengths to train ourselves to avoid these kinds of lapses. Often, doing so is central to one’s job description. Consider the accountant who makes an oversight that gets a coworker fired. I argued in chapter four that such a blunder need not reflect a lack of effort or care—sometimes mistakes just happen. But this is hardly the standard case; it is easy enough to imagine that such an error could occur as a result of laziness, haste, or apathy. Because of the potential implications of such oversights, and the fact that focus, diligence, and double-checking can help to decrease the frequency of their occurrence, it makes sense that we do not excuse “mere overlooking.” Similarly, there are reasonable steps one could take to increase the chances of remembering to discharge a duty, and because we want to minimize the number of memory failures resulting from a lack of effort or care, “I just forgot” does not count as a good excuse. Or consider Blair Walsh, the unfortunate field goal kicker. The expectation in play was that he make the kick, not that he try his hardest; converting kicks is his only task. By practicing and preparing for games, kickers can minimize their misses. For this reason, “I just missed” is not an excuse; the “don’t miss easy kicks” rule does important ethical work.

Of course, it is important to the kicker, accountant, and bus stop pickup cases that one could not have reasonably expected the blamed agents to will themselves to do better, nor to care any more, than they actually did. The quality of their wills is exemplary. Nevertheless, they are liable to blame because they failed to meet reasonable expectations without a valid excuse.

This sketch is meant to provide at least a preliminary explanation of the intuition that spasms, for example, and sudden alien interventions, function as excuses in ways that non-disclosive forgettings, lapses, and errors do not. But drawing that picture is only half of the task. The question we set out to answer was whether an alternative set of rules, one that was more fine-grained and prohibited blame for non-expressive action, would be a better set of rules than the one I endorse.

Again, I must stress that such a question is difficult to evaluate. The kinds of considerations that can make a set of rules better or worse than some alternative set are various and hard to measure. Still, we
can make some progress. For the following reasons, I think we should prefer the imprecise practice to the more exacting alternative.

First, there are the considerations of deterrence and encouragement I raised in my elucidation and defense of the idealized version of our existing excuse-rules. Because lapses, failures of memory, mistakes, and errors are often the result of an agent’s lack of good will, and because they can be minimized with work, we have good reason to set up a system that encourages credible effort. Would the fine-grained practice be able to do this? I’m skeptical. It is often difficult to tell how hard one is trying, and success or failure in a task can serve, practically, as a necessary barometer for making approximate judgments. Moral rules, and the excuses that complicate them, must be general. Their action-guiding utility depends on their being broad enough to cover a wide range of circumstances.

It is also worth considering the potential for abuse of the granular excuse-rule. What is to stop anyone who, after falling short as a result of recklessness or even malice, wishes to appeal to it? There is often no way to verify whether an erring agent made a reasonable effort to avoid his or her mistake. How would we go about confirming whether such an agent really deserved to be let off the hook? Could agents themselves even be sure?

Having one’s blameworthiness hinge on success or failure may also have the effect of strengthening one’s resolve to try. Sometimes we do not know how hard we can push ourselves until we are under pressure. The injunction to “do your best” will rarely succeed in actually getting the best out of someone who does not know the depths of his reserves. Strict standards such as “add the numbers up correctly,” or “keep your promise,” may be more motivating.

Fundamentally, we should remember that we create and enforce rules, moral and otherwise, to perform practical functions. We care about their ability to help us to get things done. Ultimately, we desire, and even require, that promises are kept, accounts are balanced, surgeries succeed, and kicks are made. Effort is nice, but we reasonably prefer even lucky completion of a required task to valiant but futile struggle.
At this point, one may grant that the imprecise excuse-rules I have been trying to vindicate have practical advantages, but that they must nevertheless be rejected on moral grounds. How should we respond to the claim, for example, that a rule is cruel because it subjects people to the unpleasantness of blame on the basis of non-expressive behavior?

In order to decide whether the rules I favor are cruel, we need to consider the kinds of “unpleasant” attitudes they sanctions in response to non-expressive rule-breaking. To that end, recall that blame is primarily a means of getting a target’s attention and of prompting apology. As I stressed in section II, blame need not involve any moral criticism or critique of the target’s rational activity. Moreover, standard episodes of blame for non-expressive failure will be short-lived. The blamed agent will know that he has violated a rule he identifies with and holds others to, and which governs a relationship he cares about. As a result, he will likely experience the pain of guilt, a moral feeling that arises when one is aware he has violated a norm of conduct and hurt someone. This feeling, in turn, may move him to apologize. If he is not so moved, blame may be appropriate as a means of calling his attention to the fact that his lapse, failure, or mistake has caused harm to another. It seems a stretch to call such rules, which would permit this kind of moral interaction, “cruel,” especially given how natural the feelings involved seem to be. Just imagine forgetting to pick a friend up from the bus stop and stranding her in the cold!

I suspect that the impulse to reject the excuse-rules I have been defending stems from the worry I have been trying to dissolve since chapter three: namely, the idea that blame should perfectly track one’s level of controlled expression of ill will. If one is in the grips of this impulse, then any system of reactive feedback that sanctions any difference in reactive blameworthiness as a result of moral luck, indeed, any system of response that could accomplish what I see as blame’s central goal—the restoration of status and the healing of social rupture in the aftermath of wrongdoing—will seem “cruel.” But blame, at least as we know it in real life, responds, and should respond to, more than the controlled expression of one’s will. One might think of this project as an extended argument for this thesis.
V: CONCLUSION: WIDE AGENCY AND THE SHALLOW SELF

Early in this essay, I argued that accounts of attributability are beholden to ethical theories of the justification of the reactive attitudes, and not the other way around. One might reasonably wonder, then, what kind of conception of attributability my understanding of the ethics of blame generates.

Because my understanding of the ethics of blame is rule-based, the responsible self that emerges will be the self that is subject to social rules. On this basis, we can rule out as non-responsible anyone who cannot be part of a rule-governed moral community. Animals, infants, psychopaths, and others who cannot understand and/or comply with moral demands, for example, do not, per my view, have “responsible selves.” This question of the bounds of responsible selfhood in general, of what membership in a moral community requires, is an important one, but I will not address it any further here.

The focus of this work has been the conditions of responsibility for individual actions. According to my theory, how must a member of a moral community be connected to her action such that she may be liable to blame on its basis? I have argued that blameworthiness requires the violation of a good social rule. As such, questions of attributability will be settled by the kind of behavior required for rules violations, and moral rules, as we have seen, sometimes respond to the wide agency of shallow selves. Still, it isn’t as though we can make no progress; surely the blameworthy actions of others are not attributable to me!

The question of attributability is not insignificant, but it should be divorced from the conventional search for the conditions of moral responsibility. If my analysis of blameworthiness is correct, all we need from a notion of attributability is a way to distinguish between one person and another. Responsibility theory has traditionally been more ambitious. Likely because of the residual effects of the concern with the metaphysics of free will, responsibility theory has understood blameworthiness roughly as one’s liability to punishment by an all-knowing god, rather than as a fundamentally interpersonal construct.

148 Contrast this understanding of the self with the self that could be the proper target of God’s eternal punishment. That “self” is deep, expressed only in the controlled action of a radically free will, and as narrow as, to use Nagel’s phrase, an “extensionless point.” See: Nagel, T. (1979). Mortal Questions. Cambridge: Cambridge University Press, 35.
shaped by the norms of personal relationships. As a result, it has felt a need to connect blameworthy action to the soul, a requirement we can ease if we embrace the social understanding of blame I have endorsed here.

In fact, the desire to interpret attributability as anything beyond the commonsense notion expressed by phrases such as “I did it,” “I caused it,” or “It was me,” is what got responsibility theory into trouble, causing theorists to reject truths about action ownership that seem obvious to everyone but responsibility theorists. Was there ever a doubt in any non-philosopher’s mind, for example, that Blair Walsh was the one who missed that consequential kick, or that it was the protagonist of Force Majeure who ditched his family? Even Williams’s unlucky lorry driver knows that running over the child is attributable to him in some meaningful sense. Whatever relationship between agent and action is required for the lorry driver to be justified in feeling that he is connected to that horrible episode in a way that others are not is enough for blame. Once that threshold, however it is ultimately spelled out, is crossed, all of the interesting questions will fall where they belong, squarely in the domain of the ethics of reactive response.
REFERENCES


