Title
How Blame Functions: Essays on Blame, Responsibility, and Moral Emotions

Permalink
https://escholarship.org/uc/item/01q141n3

Author
Wang, Tinghao

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

How Blame Functions: Essays on Blame, Responsibility, and Moral Emotions

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

in

Philosophy

by

Tinghao Wang

Committee in charge:

    Professor David O. Brink, Chair
    Professor Craig R. M. McKenzie
    Professor Dana Kay Nelkin
    Professor Manuel Rogelio Vargas
    Professor Monique L. Wonderly

2022

The dissertation of Tinghao Wang is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I want to thank many people for their help with the completion of this dissertation. My gratitude first goes to the chair of my committee, David Brink, for his constant feedback and guidance throughout the different stages of this work, from the most miniature details of the argumentation to the most general framings of the project. I also owe so many thanks to other members of my committee: Dana Nelkin, Manuel Vargas, Monique Wonderly, and Craig McKenzie. Dana Nelkin has given so much written and verbal feedback for all seven chapters; Chapter 6 in particular started with a half-baked idea about shame and guilt, and it's Dana who has encouraged me to further develop the idea and introduced me to a relevant cutting-edge debate. Manuel Vargas has given me some of the most useful big-picture comments I have ever had, and he suggested the current order in which I present the different chapters in this final version. Monique Wonderly has given constructive feedback especially from the perspective of the philosophy of emotions, and she encouraged me to work out the argument against the communication argument in Chapter 2. Craig McKenzie has given invaluable feedback from the perspective of psychology, and he has been a continuing source of support for the interdisciplinary questions relevant to my work. To you all, I cannot thank enough.

Many philosophers, colleagues, and friends have given comments to earlier drafts of parts of this dissertation. Among them, I want to thank especially Craig Agule, William Albuquerque, Santiago Amaya, Henry Argetsinger, Richard Arneson, Chris Atkinson, David Beglin, Gunnar Björnsson, Mara Bollard, Andreas Brekke Carlsson, Rosalind Chaplin, Aaron Chipp-Miller, Kathleen Connelly, Cory Davia, Derick Hughes, Drew Johnson, Kirstine La Cour, Coleen Macnamara, Joseph Martinez, Michael McKenna, Leonardo Moauro, Douglas Portmore, Daniel Ranweiler, David Shoemaker, Ahmed Siddiqi, Angela Smith, Danieal Telech, Marshall Thompson, Bas Tönissen, and Daniel Weltman.

Thanks to my parents, for their unconditional support for my career. And to Ying Liu, for all her love and support.

Chapter 2, in part, has been accepted for publication (as "The Communication Argument and the Pluralist Challenge") in Canadian Journal of Philosophy, 2021, Wang, Shawn Tinghao. The dissertation/thesis author was the primary investigator and author of this paper.

Chapter 3, in part, is currently being prepared for submission for publication of the material (as "Blame: A Pluralist Function-Based Approach," by Wang, Shawn Tinghao). The dissertation/thesis author was the primary investigator and author of this paper.

Chapter 4, in part, is currently being prepared for submission for publication of the material (as "Blame Conflict," by Wang, Shawn Tinghao). The dissertation/thesis author was the primary investigator and author of this paper.

Chapter 5, in part, has been accepted for publication (as "Response-Dependence in Responsibility: A Granularity Challenge") and may appear in American Philosophical Quarterly, 2022, Wang, Shawn Tinghao. The dissertation/thesis author was the primary investigator and author of this paper.

Chapter 6, in part, has been accepted for publication (as "Shame and the Scope of Moral Accountability") in The Philosophical Quarterly, 2021, Wang, Shawn Tinghao. The dissertation/thesis author was the primary investigator and author of this paper.

VITA

2014	B. A. in Philosophy, Renmin University of China, Beijing, China
2016	M. Phil. in Philosophy, Lingnan University, Hong Kong
2022	Ph. D. in Philosophy, University of California San Diego

ABSTRACT OF THE DISSERTATION

How Blame Functions: Essays on Blame, Responsibility, and Moral Emotions

by

Tinghao Wang

Doctor of Philosophy in Philosophy

University of California San Diego, 2022

Professor David O. Brink, Chair

What it is for us to blame an individual, and when is our blame appropriate? While moral psychologists have proposed a variety of theoretical approaches, they only recently started to systematically investigate the *function* of blaming practices and its bearing on questions about blame's nature and normativity. In developing their function-based approaches, moral psychologists have not, however, paid much attention to the possibility that blame may have multiple functions.

My project aims to fill in this gap by developing a pluralist function-based approach to both the nature and the ethics of blame. Chapter 1 introduces the backgrounds and themes of the project. Chapter 2 directly engages with and challenges what I take to be a monist aim-based

or function-based approach—the view that the aim of blame involves moral communication. Chapter 3 develops my pluralist function-based approach to the nature of blame. I suggest that blame is defined by a cluster of functions including protest, communication, and signaling. I contend that these functions are not merely accidentally associated with each other; instead, they, when realized together, have a distinct value of contributing to various first-order normative aims in an efficient and non-alienating way. Chapter 4 applies this approach to a kind of practical conflict about blame. Sometimes we face conflicting considerations about whether to blame because, I argue, blame cannot fulfill its multiple functions at the same time. In Chapters 5 and 6, I turn to examine questions about the role of reactive attitudes in theorizing moral responsibility. I argue against an influential "response-dependence" view that grounds facts about responsibility in facts about reactive attitudes; and I defend an inclusive view about the scope of reactive attitudes, according to which the controversial attitude of shame should count as a reactive attitude. The conclusion is that a pluralist function-based approach has various theoretical advantages, and reactive attitudes should play an important but limited role in this approach.

Chapter 1

Introduction

Blame is a central, if not the most central, response that we tend to have when we see a person doing a morally questionable thing. It is surely not the only possible response. Sometimes we have harsher, and "colder," responses, like a punishment, a threat, or a warning—e.g., "I won't let it go if you do this again!" Sometimes we have milder, and maybe "warmer," responses compared to blame, such as feeling and expressing disappointment, frustration, or sadness—e.g., "it makes me really sad you did this." Blame can *accompany* these nearby practices. For example, punishment may often involve blame but goes further; and sadness may often be an aftereffect of blame. However, blame appears to be a *distinct* phenomenon that stands in between, and shares some similarities with, these harsher or milder alternatives. On one hand, blame does not exactly amount to a punishment, but it often inflicts unwanted guilty and painful feelings on its target, and in this fairly thin sense it may be said to count as a sanction. On the other hand, blame goes beyond mere disappointment or sadness, but it often also involves having or expressing a negative moral emotion, such as resentment and indignation.

One philosophical question about blame, then, is what makes blame a distinct response to norm violations, setting it apart from the harsher or milder alternatives. Issues about blame have concerned contemporary analytical philosophers since the mid-to-late 20th-century, driven

by the problem of free will. The problem is that, if it turned out that we lack free will, and are not morally responsible, due to causal determinism, would it ever be appropriate for us to blame someone? In the last decade or so, there has been another surge of literature concerned with blame, though the recent authors are generally (but not always) more inclined to examine the nature and ethics of blame itself, independent from the problem of free will. Many differing accounts about the distinctiveness of blame have been offered.

A partial list is as follows. Note that some of these views are more concerned with the nature of blame (what blame is), whereas others are more concerned with the propriety condition of blame (what makes blame appropriate). I put them together to showcase the variety of approaches to identifying what is distinctive about blame. One may also endorse the conjunction or disjunction of more than one of the following claims, as can been seen by the fact that some authors appear in more than one place on the list.

1. Blame consists of a belief or judgment that its target is morally blameworthy for something.[1]
2. Blame consists of a belief-desire pair, including a belief that its target has done something wrong and a desire that she did not do it.[2]
3. Blame consists of an emotional attitude such as resentment, indignation, and guilt.[3]
4. Blame is primarily an outward behavior or speech act, such as a complaint or a condemnation, that expresses one's negative attitude about something done by its target.[4]
5. Blame involves a demand for reasonable regard.[5]
6. Blame involves a desire for payback or retribution.[6]
7. Appropriate blame is deserved by the individual who receives it.[7]
8. Blame involves taking one's relationship with another person to be impaired.[8]

---

[1] See, e.g., Hieronymi (2004), Brink & Nelkin (2022), and Portmore (2022).
[2] See, e.g., Sher (2006).
[3] See, e.g., Wallace (1994, 2011), Wolf (2011), Tognazzini (2013), Pickard (2013), Cogley (2013), Rosen (2015), and Menges (2017a).
[4] See, e.g., McKenna (2012, 2014) and Fricker (2016).
[5] See, e.g., Watson (2004), Darwall (2006), and Shoemaker (2007).
[6] See, e.g., Strawson (1962) and Nussbaum (2016).
[7] See, e.g., Clarke (2013), Pereboom (2014), and Nelkin (2016a).
[8] See, e.g., Scanlon (2008).

9. Blame involves protesting the moral claim implicit in another person's conduct.[9]

10. Blame involves a communicative/conversational process between the interlocutors.[10]

11. Blame involves issuing a costly signal that one is committed to the enforcement of a norm.[11]

12. Blame is a mechanism of social regulation that aims to ensure that people act according to norms.[12]

13. Blame derives its justification from the way it improves our moral capacities and cultivates better moral agents.[13]

14. Blame distinctively involves a "proleptic" mechanism: it causes people to recognize new reasons by treating them as if they already recognize them.[14]

15. Appropriate blame tracks unacceptable social imbalances in the social strength between individuals, with a wish for the restoration of social equilibrium.[15]

A rough divide between these views is in terms of whether they mark out what is distinctive about blame in terms of its *constitution* or in terms of its *function*. 1–7 focus more on its constitution. On these views, we understand what is special about blame by asking questions like: Is blame constituted by psychological states or actions? If former, then what psychological states constitute blame? What are the representational contents of these states? By contrast, 8–15 focus more on the function of blame. On these views, we should be primarily concerned with questions like: What is the function of blame? How does this make blame different from other attitudes and practices?[16]

---

[9]See, e.g., Boxill (1976), Hieronymi (2001), Talbert (2012), and Smith (2013).

[10]See, e.g., Watson (2004, 2011), Darwall (2006), Shoemaker (2007), McKenna (2012), Macnamara (2015a, 2015b), and Fricker (2016, 2018).

[11]See, e.g., McGeer (2013), Kogelmann & Wallace (2018), and Shoemaker & Vargas (2021).

[12]See, e.g., Nowell-Smith (1948), Smart (1961), and Arneson (2003).

[13]See, e.g., Vargas (2013, 2021) and McGeer (2013).

[14]See, e.g., Williams (1995), Fricker (2016, 2018), Bagley (2017), and McGeer (2019).

[15]See, e.g., Reis-Dennis (2021).

[16]This is, admittedly, a very rough divide of the landscape of the literature, and there are various nuanced versions when it comes to each of these views. For example, some function-based views about blame are consistent with there also being content-based constraints (e.g., Shoemaker & Vargas 2021; Reis-Dennis 2021). The desert-based view (7) is often about blame's content—blame involves a belief that someone deserves negative treatment (see, e.g., Portmore 2019a, 2019b); but sometimes desert is meant to be about not blame's content but rather its propriety condition (see, e.g., Carlsson 2019). The relationship-impairment view (8) is another example that, I think, can be reasonably interpreted as either constitution-based or function-based.

The broad goal of this dissertation is to develop what I take to be the best function-based approach to blame. On my view, the key to identifying what is special about blame involves examining the functions of the practice. Blame functions differently from punishment, threat, warning, disappointment, frustration, and sadness. The distinct functional profile of blame, I shall argue, is what we need in order to decide who are the proper targets of blame, to define what blame is, and to understand an important kind of practical conflict about blame.

The central feature of my account is that I will combine a function-based approach with a pluralist approach to blame. Theorists who have defended views among 8–15 sometimes use function-talk, but they still tend to assume or write as if blame has one single, unifying function. Indeed, it may seem that function-based approaches will not work unless we can find this single function. I worry that this sells the function-based approach short. Instead, I will motivate and defend a pluralist function-based approach that understands blame in terms of a number of different but interrelated functions.

Another feature of my account concerns the role that reactive attitudes play. What are reactive attitudes? They include emotions such as guilt, resentment, and indignation; they often are responses—thus "reactive"—to people's actions that manifest a good, poor, or indifferent quality of will; and they are widely thought of as being closely tied to, in one way or another, our practices involving holding other people, and ourselves, morally responsible. I argue that reactive attitudes have a crucial but limited role to play in the metaphysics of moral responsibility; and I also argue for a broad, inclusivist view on the scope of the reactive attitudes. Both views— a restrictionist view about the metaphysical role of reactive attitudes and an inclusivist view about the scope of such emotions—are meant to enrich and relate to the pluralist function-based framework.

In what follows, I will introduce and motivate these three main features of this dissertation: the appeal to blaming practices and blame's functions (§1.1), the pluralist element (§1.2), and the role of reactive attitudes (§1.3). I then end with a brief summary of different chapters (§1.4).

1.1 Blame, Practice, and Functions

Function-based approaches study the distinctive feature of blame in terms of its function. They either do so by suggesting that blame is only *fitting* or *appropriate* when it fits its function, or, as perhaps a more radical claim, suggesting that blame just *is* what plays a certain functional role. Either way, the method of the investigation is to start from how blame functions, and to proceed to gain some insights about the nature and ethics of blame. But why should we start with functions? I will offer three motivations: one based on a general metaphilosophical view, and two based on more specific observations about blame.

1. The general metaphilosophical motivation is the idea that philosophical investigation of a practice *X* could benefit from starting with studying how people actually engage in the practice *X*. To explain, compare this practice-driven methodology with what is sometimes called the "standard" intuitive-driven philosophical methodology. The standard intuition-driven methodology, or perhaps a caricature of it, is as follows: we start from examining people's case intuitions and general intuitions about *X*, and try to find a theory of *X* that does the best job of fitting all the intuitive data. By contrast, the primary object of concern for a practice-based driven methodology is how we actually engage in the practice *X*, instead of how we think or intuit about *X*—in a nutshell, the focus is on *what we do*, instead of *what we think*. The idea is that examining our practices about *X* would produce a distinctive set of data points that can enrich our theorizing of the phenomena. Intuitions still play a crucial role in this process, but they only do so by supplementing and providing evidence for facts about our practices.[17] If such a practice-based methodology is on the right track, then the function-based approach to blame constitutes an appealing way of focusing on a particular kind of facts about our practices, through the examination of the functional roles of blame.

---

[17] Another possibility is that our practices produce theoretical data points by revealing our *tacit* beliefs, and it is only for this reason that practices matter for our theorization. This is one way of understanding Rawls's (1971) method of reflective equilibrium.

2. There are two specific reasons why such a practice-based methodology is especially appealing when it comes to blame. The first reason comes from an observation by Strawson (1962), through his claim that blame involves a reactive attitude that "neither calls for, nor permits, an external 'rational' justification" and of which "questions of justification are internal to the structure or relate to modifications internal to it" (64). It is notoriously hard to spell out what exactly the observation is, but it has apparently moved many theorists to endorse a, broadly speaking, "Strawsonian" approach to blame and moral responsibility. The basic idea is that we should adopt a practice-based methodology about blame not because of its general appeal, but because this is the *only* methodology available to us when it comes to blame. We cannot have a theory of blame that departs from our blaming practices—even if we could, the resulting practice would not be psychologically possible for us to live with. This very general idea has been given different variations and labels: a "Strawsonian,"[18] "naturalist,"[19] or "practice-based"[20] theory; a "naturalistic strategy";[21] or "Strawson's reversal."[22] One possible way to put the appeal of the idea is as follows. Blame, as a type of practice, is pervasive and personal. It is pervasive in that we blame a lot; it is not a practice that we only rarely engage in. It is also personal in that we often have a personally invested interest when we blame. The fact that blame is so pervasive and personal makes it such an important part of our moral life; it is fairly hard to conceive how we can live our life without it. Therefore, whatever theory of blame we have had better reflect what blame really is like in our life. If our theory fails to match our practice well, then we would likely end up with a practice that is not at all pervasive, because it diverges from our ordinary practices, and not at all personal, because it diverges from our ordinary, often personal, experiences when we blame. And to ensure that we do not end up

---

[18]E.g., Watson (2004).
[19]E.g., Campbell (2017).
[20]E.g., Vargas (2004).
[21]E.g., Russell (1992).
[22]E.g., Todd (2016).

with such a theory, we had better make our practices about blame the starting point of our theorization. Again, a function-based approach picks out a particular kind of facts that a practice-centered approach should examine: facts about the functional roles of our blaming practices.

3. It is fair to say that this Strawsonian point is what has moved many theorists to endorse a practice-centered or function-centered methodology. There is, however, another reason to do so that has gone largely unnoticed, which provides some direct motivation for adopting a function-centered approach. This is the observation that the *second-order* nature of blame makes a function-based methodology especially appealing. Let me explain. It is quite plausible that blame is about norms. If there were no norms in the world, we probably would not have blaming practices. But blame is also about norms in a second-order way: we blame *S* for her action *A* when *A* violates a norm *N*. It is "second-order" in the sense that the connection between blame and a norm is always mediated by an action *A*. By contrast, the action *A* has a first-order connection to the norm *N*: *A* either follows or violates *N*. But why do we then need this second-order activity called blame? Why don't the first-order norm-following and norm-violating activities exhaust our norm-related practices? The most natural answer to these questions is that there is something distinct that blame *does* with regard to the more basic first-order moral practices. One way to spell this out is a functionalist proposal: blame's distinct value comes from its *functional role* in contributing to the more basic first-order moral system.

To illustrate this last idea, consider the relationship between blame and punishment. Blame is not yet punishment. However, when it comes to this second-order nature, there are interesting similarities. Punishment, like blame, also connects to some first-order norms—legal norms—in a second-order way—we punish people who violate the first-order legal norms. Punishment, through its functional role, has the distinct value of promoting these first-order legal norms. If we

aim to arrive at a good theory of what punishment is and when punishment is appropriate, we probably want to take as starting points the functional roles of punishing practices. It is the same methodological orientation that function-based approaches aim to advocate about blame.

One may find a Strawsonian, practice-centered approach to blame appealing, but be skeptical about the idea that we should examine the function of the practice in particular. Why not just look at, for instance, the general pattern of people's attitudes and behaviors when they blame? Why do we need to introduce a functional or teleological element? A goal of this dissertation is to answer these questions, by developing a function-based framework that, I shall argue, has various theoretical advantages. I hope to arrive at an adequate answer by the end of the work. But one reason to even pursue such an approach to begin with, I think, is the potential explanatory value of function-claims. One issue with a practice-based methodology is that it seems to lack explanatory value. Just because we have a descriptive theory of what we do, it might be suggested, this does not mean that it explains why, as a normative matter, we ought to act according to such a theory. This is why I am pessimistic about a practice-based approach that simply starts from what we do in each concrete situation, or even the general patterns of how we think and act. Instead, I think that we have to look at something deeper—What is the function of this practice? Does our doing things in this way serve any theoretically or practically important purposes?

I find it promising that these function-claims are indeed deeper claims about our practices— they do carry with some explanatory depth. One reason to believe this is that function-claims have been granted an explanatory role in many other scientific and philosophical fields. One such field is biology. Functional statements are commonly used to explain biological phenomena, and a huge debate in the philosophy of biology concerns the nature of functional explanation. Various accounts have been proposed, including deductive-nomological accounts (Nagel 1961), causal role accounts (Cummins 1975), evolutionary accounts (Neander 1991b), and design-oriented accounts (Kitcher 1993). Despite their disagreements on what functional explanation is, and how unique it is as a kind of scientific explanation, these theorists all agree that functional explanation

is a real thing. As another example, functionalist approaches to mental states—the idea that some, or even all, psychological states just *are* what have a certain functional profile—have been influential in the philosophy of mind. The mental state of being in pain, for example, does not seem to depend on a particular physical or chemical constitution, since a variety of creatures—humans, birds, snakes—can experience pain but have radically different physical and chemical constitutions. According to functionalism about pain, this mental state is just whatever state that has a certain function (e.g., signaling potential damage and causing beliefs of avoidance), which can be realized in different physical and chemical forms for different creatures. Some philosophers of mind (e.g., Fodor 1968) then take a further step: functionalism is both a theory of what psychological states are and a theory of what *psychological explanation* is. When we use pain to explain our thoughts and behaviors, then, we in effect rely on pain's function (e.g., to signal potential damage and to cause beliefs of avoidance) in our explanation.

I am thus inclined to believe that function-claims about blame will turn out to have some explanatory values as well. The open question, however, is what kind of explanatory values they have, and whether they are fitting for the kind of job that a theory of blame is meant to accomplish. Part of the aim of this dissertation is to explore this open question by developing what I take to be the most promising function-based theoretical framework.

1.2   Pluralism about Blame and Moral Responsibility

As I have mentioned earlier, my function-based approach is *pluralist*. Specifically, I will argue that blame has at least three definitive functions: protest, communication, and signaling. I am also committed to a widely accepted connection between blame and moral responsibility: it is appropriate to blame *S* for her action *A* only if (i) *S*'s action *A* violates some norm, and (ii) *S* is morally responsible for *A*. This makes my account similar to, and continuous with, other pluralist (or "hybrid," "disjunctive," "combination") views about blame and/or about moral responsibility

9

in the literature. In this section, I will survey some existing pluralist views and compare them to the approach that I will develop.

Watson (1996) distinguishes between two kinds of blame and two kinds of moral responsibility. On one hand, there is *aretaic blame* associated with moral responsibility in the *attributability* sense. On Watson's view, one is morally responsible for an action in the attributability sense just in case that the action "was attributable to the agent," and to blame in the aretaic or attributability sense is just to "attribute something to a fault in the agent" (1996, 230–1). On the other hand, there is blame in the sense of *holding someone accountable*, associated with moral responsibility in the *accountability* sense. According to Watson, one is morally accountable for an action just in case that one has proper control over the action. And to blame in the accountability sense is to hold someone accountable, and to impose an informal kind of sanction. Appropriate blame, in this sense, must be deserved by its target.

Though Watson's distinction between attributability and accountability has been the most influential, there are other proposals that distinguish between two or three kinds of moral responsibility as well, such as Zimmerman's (1988) distinction between "appraisability" and "liability"; Scanlon's (1998, 2015) distinction between "attributability" (what he later calls "moral reaction responsibility") and "substantive responsibility"; and Shoemaker's (2015) distinction between "attributability," "answerability," and "accountability." Fischer and Tognazzini (2011) explore an even more fine-grained distinction that involves distinguishing different kinds of attributability and different kinds of accountability as well.

I find Watson's taxonomy a useful way to distinguish between two kinds of practices. But I am skeptical if what he calls "aretaic blame" is genuinely blame, and my focus in this dissertation will be mostly on what he calls "moral accountability." I will try to argue that, even if we focus exclusively on accountability blame, there are still reasons to favor a pluralist picture, because accountability blame plausibly plays a plural number of functions.

Another difference is that my pluralism is primarily about the *functions* that blame has,

instead of about how many things there are when we talk about blame. One advantage of this approach is that it bypasses some questions about the individuation of subtypes of blame: How many kinds of blame there are? Is blame "one thing" or "many things"? I am inclined to think that these are not the most urgent questions, and a lot of debates on these issues will turn out to be merely verbal disagreements. There are of course many different kinds of blame, and in this sense we can always distinguish between many things that "blame" refers to. But this is compatible with thinking that the umbrella term "blame" that refers to the combination of these different kinds refers to one thing. The issues that I think are more important are what *functions* blame and its subtypes serve, and how valuable these functions are. This is why I will attempt to combine a function-based approach with a pluralist approach about blame.

One existing pluralist function-based proposal is by Cogley (2013). He argues that blame has three functions: appraisal, communication, and sanction. He then suggests that these three functions correspond to three different senses in which we might say that blame is "appropriate." But Cogley fails to address questions about the value of these three functions. Why do blame have these functions? Are there any special reasons why they come together? Do they have any explanatory values? I will aim to provide a framework that does a better job at answering these questions.

1.3   The Role of Reactive Attitudes

Another major theme of this dissertation involves the claim that we ought to restrict and recalibrate the role of reactive attitudes in theorizing moral responsibility. Reactive attitudes are our emotional responses to people's actions that manifest good, poor, or indifferent quality of will. They include at least the attitudes of guilt, resentment, and indignation. On my view, these attitudes have an important but limited role in theorizing blame and responsibility. My function-based approach to blame and moral responsibility goes hand in hand with this restrictive

11

stance about the role of reactive attitudes. It is a natural consequence from the function-based approach that reactive attitudes have a crucial role to play in theorizing about blame and moral responsibility, because they are emotions that, I think, quite often count as blame due to their dispositions to fulfill various functions—protest, communication, and signaling. They are thus indeed central to our responsibility practices. However, the function-based approach also suggests that their centrality should not be overstated. Reactive attitudes have their central roles *because* they play those functions definitive of blame and responsibility practices. There can be occasions where some tokens of reactive attitudes fail to play, or even lack the disposition to play, those functions, and there can also be some tokens of non-reactive attitudes or activities that turn out to play these functions and are thereby significant for our responsibility practices.

This picture requires taking a stand on a substantial debate about the metaphysical status of reactive attitudes. This debate centers around the *priority* between facts about reactive attitudes and facts about moral responsibility.

The question here mirrors the more general question in ethics and moral psychology: Everyone agrees that moral emotions are crucial for our making moral judgments, but do moral emotions *ground* moral judgments or merely *correlate with*/provide important *evidence* for moral judgments? Sentimentalists endorse the former, stronger claim, whereas rationalists often endorse the latter, weaker claim. Similarly, the debate on the metaphysical status of reactive attitudes is between moral responsibility sentimentalists—the "response-dependence" theorists who believe that reactive attitudes ground judgments about responsibility—and moral responsibility rationalists—the "response-independence" theorists who merely believe that reactive attitudes correlate with or provide important evidence for judgments about responsibility. Because I claim that the role of reactive attitudes is derived from the functions they serve, my view is response-independent; I deny that facts about reactive attitudes are the ultimate basis from which we derive facts about moral responsibility.

Another question about reactive attitudes concerns their scope: which emotions count

as reactive attitudes? Wallace (1994) has a narrow, but fairly influential, answer, according to which only resentment, indignation, and guilt are reactive attitudes. These emotions, according to Wallace, are tied to moral responsibility because they involve a "demand," but emotions like shame, disgust, and contempt do not. Darwall is another philosopher who has suggested a fairly narrow view on the scope of reactive attitudes. He contends that reactive attitudes need to be both "second-personal" and "accountability-seeking" (Darwall 2015); while guilt has both features, shame is excluded because it's "third-personal" instead of "second-personal" (Darwall 2006, 71). In other works, Darwall includes a wider range of attitudes, including love (2016), trust (2017), and gratitude (2020), as "second-personal" and "reciprocation-seeking"; he believes that they are crucial interpersonal attitudes "of the heart," but he still denies that they are reactive attitudes in the accountability-seeking sense (see 2015). In contrast, some theorists endorse a fairly broad view about what counts as reactive attitudes. In Strawson's seminal paper on reactive attitudes, he includes a long list of emotions including "gratitude, resentment, forgiveness, love, and hurt feelings" (1962, 48); and "feeling bound or obliged (the 'sense of obligation'): feeling compunction; feeling guilty or remorseful or at least responsible; and the more complicated phenomenon of shame" (1962, 57). And while most philosophers now would deny that *all* these attitudes are reactive attitudes, many suggest that some of them are, in addition to the three as identified in the narrower view (resentment, indignation, and guilt). Some examples are trust (Helm 2014), contempt (Mason 2018), hurt feelings (Shoemaker 2019), and even spite (Thomason 2020).

I am committed to a broad view about the scope of reactive attitudes. Emotions like shame, for example, should count as reactive attitudes in the accountability-seeking sense. This is continuous with the pluralist function-based proposal. I have mentioned that, even when we narrow down on moral accountability, our blaming practices still play multiple functions. It is thus unsurprising that we find different kinds of blaming practices—guilt and shame in the case of self-blame—that are better suited for playing different functions, though both would count as

blame in the sense tied to accountability.

1.4   The Plan of the Work

In Chapter 2, I engage with an influential line of argument: communicative capacities are necessary for morally responsible agency because blame *aims at* a distinctive kind of moral communication. I contend that existing versions of the argument face a "pluralist challenge": they do not seem to sit well with the plausible view that blame has multiple aims. I then examine three possible rejoinders to the challenge, suggesting that a context-specific function-based approach constitutes the most promising modification of the communication argument. I contend that this is the best way to modify the communicative theorist's argument; it retains to a great extent the normative significance of moral communication in blame, and it fits better with the pluralist view about blame's aims and functions.

In Chapter 3, I propose a pluralist function-based approach to defining what blame is. Existing function-based approaches appear to assume that blame has only one single, unifying function, but I argue that a pluralist approach fits better with our best accounts of what functions are. Another feature of the existing approaches is that they tend to define the blame-type without offering a definition of blame-token. I think this is unsatisfying. Instead, I offer both a definition of blame-type and a definition of blame-token. The key move is the suggestion that blame is defined, as a type, by three functions: protest, communication, and signaling. Individual tokens of blame are then defined by their dispositions to fulfill enough of these functions.

In Chapter 4, I apply the function-based approach to understanding a kind of practical conflict, where we are faced with conflicting considerations about whether to blame. What if, for example, blame would fulfill its protest and signaling function but cannot fulfill its communicative function? I argue that this kind of practical conflict happens because of conflicts between blame's different functions. Though blame is designed to fulfill multiple functions at the same time,

these are the non-ideal cases where fulfilling one function would in fact undermine another of blame's functions. I then discuss various kinds of factors that we should take into consideration in deciding whether to blame, including factors about the general significance between the conflicting functions, their context-specific significance, and if there are any significant higher-order functions that blame or withholding blame would be likely to fulfill.

In Chapter 5, I turn to discuss the metaphysical role of reactive attitudes in theorizing moral responsibility. I argue that the response-dependence view, according to which we are morally responsible in virtue of our being appropriate targets of reactive attitudes, faces a serious challenge. The challenge is about the granularity of reactive attitudes. As emotional attitudes, they are too coarse-grained to ground facts about the more fine-grained facts about moral responsibility. The upshot is that reactive attitudes may be a crucial part of the metaphysical explanation of facts about moral responsibility, but further theoretical resources are necessary to make sense of why we can be morally responsible to a fine degree.

In Chapter 6, I discuss the issue about the scope of reactive attitudes, by focusing on the controversial attitude of shame. I argue that shame, or at least a kind of shame, should count as a reactive attitude. In making the argument, I appeal to a fairly broad view on the scope of moral responsibility, according to which we are morally responsible for not just our actions, but also our attitudes, character, and self. But shame, rather than guilt, better captures our corresponding emotional responses when it comes to responsibility for attitudes, character, and self.

Chapter 2

The Communication Argument and the Pluralist Challenge

2.1 Introduction

In an influential remark, Watson writes that,

In a certain sense, blaming and praising those with diminished moral understanding loses its "point"... The reactive attitudes are incipient forms of communication... the most appropriate and direct expression of resentment is to address the other with a complaint and a demand. Being a child exempts, when it does, not because expressing resentment has no desirable effects; in fact, it often does. Rather, the reactive attitudes lose their point as forms of moral address. (2004, 230–231)

This passage has been frequently cited approvingly, and the argument in this passage— from how blame would lose its "point" to why we should exempt individuals—has received much discussion in the literature (e.g., Darwall 2006; Shoemaker 2007; Watson 2011; McKenna 2012; Talbert 2012; Smith 2013; Macnamara 2015a, 2015b; Beglin 2020). But what does it exactly mean to say that blaming attitudes or behaviors "lose their point as forms of moral address"? I take it that the most natural construal is a teleological one: blame loses its point just in case blame does not *fit its aim*. The argument, then, is that communicative capacities (e.g., the capacity

of moral understanding) are necessary for morally responsible agency because blame *aims at* a distinctive kind of moral communication. Following Macnamara's terminology (2015a, 212), I will refer to this line of reasoning as the "communication argument."

Many have endorsed the communication argument. For example, McKenna states that communicative capacities are necessary for responsible agency, because "blame only has a point if the one toward whom it is directed is able to understand it as an expression of our moral regard for, and expectations in response to, what the person has done" (2012, 109). As another example, Shoemaker puts it that psychopaths and moral fetishists are not morally responsible agents because they cannot care about the basic demand to be communicated, and thus blaming them would be "particularly pointless" (2007, 85). In his later work, Watson himself draws the same conclusion about psychopaths by contending that the "*telos*" of blaming practices is "the prospect of codeliberation and reconciliation," but when it comes to psychopaths "this hope is forlorn" (2011, 323). And Macnamara (2015a, 2015b) offers a functional account of the communication argument, according to which reactive attitudes are "communicative entities" in the sense that they "have the etiological function of eliciting sincere acknowledgment of fault from the wrongdoer" (Macnamara 2015a, 222).

The goal of this chapter is to present a novel challenge against the communication argument, based on the plausible pluralist view that blame has more than one aim. Though various theorists have mentioned or briefly discussed the possibility that blame has multiple aims, its potential threat to the communication argument has not been sufficiently discussed. I will argue that there is a serious difficulty in configuring the communication argument under a pluralist picture about blame's aims. I then discuss three possible strategies to respond to the difficulty: a conjunctive approach, a disjunctive approach, and a context-specific approach. I conclude by suggesting that the context-specific approach, when combined with a functional account of the communication argument, offers the most promising way out of the pluralist challenge.

Before I proceed, it is worth noting that the communication argument is not the *only*

argument for the alleged centrality of communicative capacities to responsible agency. Communicative theories of responsible agency consist in a broad family of views (e.g., Watson 2004, 2011; Darwall 2006; Shoemaker 2007; McKenna 2012; Macnamara 2015a, 2015b; Fricker 2016, 2018). They all share the commitment that communication is the key to understanding what it is to be a morally responsible agent, but the specific contents of and motivations behind these views vary significantly. For example, Darwall's version of the theory is part and parcel of a second-personally oriented metaethical framework that other communicative accounts are usually reluctant to rely on, while McKenna's version of the theory is uniquely built upon an analogy between moral communication in blame and linguistic conversation. My interest in this chapter is not to give an overall evaluation of all those different sorts of communicative theories and potential arguments supporting each. Instead, I will address the particular issue of whether the communication argument, or a close variant thereof, is well-suited to back up the communicative theory of morally responsible agency.

Another important preliminary matter consists in the question of what blame is. I will ultimately defend a function-based account of blame (Chapter 3). For the purpose of this chapter, I will assume a more ecumenical starting point (compared to the account in Chapter 3): blame includes at least negative reactive attitudes and their outward expressions. That is, blame can be either (i) private blame, in the form of a negative emotion such as resentment, indignation, and guilt, or (ii) public blame, in the form of an outward behavior, like complaint and condemnation, that expresses a negative emotion such as resentment, indignation, and guilt. This fairly ecumenical stance on the nature of blame is neutral on the question of whether private blame or public blame is more explanatorily basic,[1] and it is compatible with various accounts about what underlying factor unifies all instances of blame.[2]

---

[1]McKenna (2012, 2016) argues that public blame is more explanatorily basic. By contrast, Driver (2016) and Menges (2017a) have provided reasons for thinking private blame is more explanatorily basic.

[2]For different accounts see, e.g., Watson (2004), Sher (2005), and Scanlon (2008).

2.2 The Communication Argument

I have described the communication argument as follows:

The Communication Argument. Communicative capacities are necessary for morally responsible agency because blame aims at a distinctive kind of moral communication.

In this section, I will more carefully unpack the communication argument by discussing (i) different views on the "distinctive kind of moral communication"; (ii) different views on the required communicative capacities; and (iii) what it means to say that one is a morally responsible agent.

According to some theorists, blame is meant to communicate a *demand* or *expectation*. This is evident in Watson's remark: "the most appropriate and direct expression of resentment is to address the other with a complaint and a demand" (2004, 230). According to Watson, blame involves a concern for the *basic demand*, which is "a *moral* demand, a demand for reasonable regard" (2004, 229). Watson seems to imply that blame, or at least the most appropriate and direct form of blame, is both meant to convey a moral demand as its *content* (the blamer is meant to communicate that she demands a reasonable degree of good will), and also meant to be in the *form* of a demand itself (her blaming attitude or behavior constitutes a demand itself, instead of other forms of communication like invitation and request).[3] In a nutshell: blame aims at communicating a demand or expectation. This is the "telos of this practice" (Watson 2011, 322).

Darwall (2006) also endorses this demand claim. However, according to Darwall, the content of the moral demand involved in blame is more complicated. Importantly, blame demands not just good will but also that others recognize the "second-personal" nature of this demand, which further requires others to recognize our *authority* to make demands. The recognition of this authority is essential when it comes to respecting a morally responsible agent as a person. As Darwall puts it, "their [reactive attitudes'] aim is to demand respect" (2006, 86) and "the

---

[3]See Macnamara (2013) and Telech (2020) for a similar distinction.

implicit aim of reactive attitudes is to make others feel our dignity" (2006, 85). Thus, according to Darwall, good will, recognition of authority, and respect are all interrelated elements involved in the content of the distinctive kind of moral demand that blame aims to communicate.

Shoemaker (2007) makes a similar point about the second-personal nature of blame's demand, but he parts ways with Darwall in taking the basic demand to involve an essentially affective dimension. Important for Shoemaker, blame is both an address of reasons and an "emotional address": it consists in "urging the wrongdoer to feel what I feel as a result of his wrongdoing and then subsequently to feel the guilt or remorse (at having caused that feeling) which I expect to motivate him to cease his wrongdoing" (2007, 91). Blame is meant to communicate a demand as "a form of emotional address"; it is analogous to asking "can't you see how you've made me feel?" (Shoemaker 2007, 100).

Though widely accepted, the demand claim has received some criticisms (e.g., Macnamara 2013; Telech 2020). The details of this dispute need not bother us here. It is enough to note that it is an open theoretical possibility for a defender of the communication argument to rely on a notion of moral address broader than that of a moral demand. As Telech puts it, "it is not clear why Strawsonians should have identified moral address with one of its varieties, namely, moral demand. Apart from demand, moral address can take the form of invitation, urging, request, and so forth" (2020, 10). Though Telech's point here is meant to apply to praise, it can also be taken as a reasonable take on the scope of moral address involved in blame. One version of the communication argument that does not rely on the demand claim has been defended by Macnamara (2015a, 2015b). Her approach is to understand the distinctive kind of moral communication in blame in terms of the *uptake* that blame is meant to receive. Macnamara puts it that uptake of blame "is understood to involve sincere acknowledgment of one's fault—most paradigmatically, feeling guilt and expressing it via apology and amends" (2015a, 215). This provides one way of understanding the communication argument while setting aside disagreements on what moral address or moral demand exactly involves, by focusing on the

"consensus that uptake requires the capacity to feel guilt" (2015a, 216). Macnamara then suggests that eliciting this emotional uptake is the aim, or, more accurately, the *function*, of blame.

The above theorists' differing views on what blame communicates correspond to their disagreements about what the relevant "communicative capacities" necessary for morally responsible agency are. Watson (2004) takes it to consist in the cognitive capacity to understand the relevant moral demand. Blaming an individual who lacks this capacity is "pointless" and "unintelligible" because it does not fit the aim of blame. Darwall concurs, but he adds that the blamer also needs to possess the capacity to "take moral demands as conclusive reasons for acting," which further involves the capacity to "hold oneself responsible" and to "determine oneself by a second-personal reason" (2006, 78). The thought is something like the following: blaming an agent who lacks the capacity to engage in the exchange of second-personal reasons does not fit the aim of blame, which is to address a demand that involves the exchange of this distinctive kind of reasons. Shoemaker (2007) agrees that a morally responsible agent needs to have the capacity to recognize and apply second-personal reasons. However, because Shoemaker thinks blame is also meant to be an emotional address, he also requires that a responsible agent has an emotional capacity: the capacity to be "moved to identifying empathy with" (2007, 101) the person who is the source of the relevant second-personal reasons. Identifying empathy requires that one is able to "feel what they [the blamers] feel in the way that they feel it" (Shoemaker 2007, 99). More accurately, Shoemaker requires that a responsible agent is either capable of *directly* empathizing with the victim of her wrongdoing, or capable of *indirectly* doing so through empathizing with an "appropriate representative" (2007, 107)—for example, her caregiver who mimics the victim of her wrongdoing. Lastly, Macnamara concludes that, because the function of blame involves evoking emotional uptakes, morally responsible agents need to possess the capacity to feel guilt.

Now we can see a similar pattern of reasoning in the above authors. They all start from a distinctive sort of moral communication that blame aims at—a reason-based demand, an emotion-based demand, or a broader form of moral address—and then conclude that, because this sort of

communication is the aim of blame, an individual who lacks the capacity to stand in the role of the addressee in the relevant communicative process is exempted from blame. After all, blaming her would not fit the aim of blame; it would be pointless. As Macnamara puts it, "it seems utterly intuitive that one is a felicitous candidate for a role only if she has the capacities necessary to perform the activities that are constitutive of the role" (2015a, 214). If we were to blame an individual who is unable to play the required role when it comes to the aim of blame, then our blame would fail to *fit this aim*. The communication argument relies on the intuitively plausible view that an instance of blame that does not fit the aim of the practice is misdirected; its target should, therefore, be exempted from our blaming practices.

Here is one way to reconstruct the steps of the communication argument in its premise-conclusion form:

The Communication Argument.

(1) Blame aims at a distinctive kind of moral communication.

(2) Blame fits this communicative aim only if directed toward those who have the relevant kind of communicative capacities.

(3) An individual is a morally responsible agent only if blaming her fits the communicative aim of the practice.

(4) Therefore, an individual is a morally responsible agent only if she has the relevant kind of communicative capacities.

Before I proceed to present my challenge against the communication argument, it is worth making a few remarks about what morally responsible agency entails. Defenders of the communication argument all agree that morally responsible agency is conceptually tied to the exemption condition of moral responsibility. An individual who lacks responsible agency is exempted from responsibility practices and exempted from blame. One contentious matter, however, is whether morally responsible agency is a context-general or context-specific concept. While the context-general view is often taken for granted by defenders of the communication argument, the context-specific view also has its proponents. For example, Vargas (2013) contends

that both empirical evidence and armchair reflection support the claim that responsible agency is likely to be a "patchy concept" such that one may have or lack responsible agency depending on the context that one is in. As Vargas puts it, capacities relevant to responsible agency can include a variety of things: "susceptibility to a nagging, reacting to a dim hope, being able to imagine the situation of another, or attending to an inarticulate, largely inchoate suspicion about things" (2013, 208–9). Given that all these elements can vary significantly across contexts, it seems unlikely that responsible agency is an entirely context-general kind.[4] Another potential complication concerns whether morally responsible agency is binary or scalar. Theorists often take for granted that responsible agency is a binary concept, but some have convincingly argued that morally responsible agency can come into degrees (e.g., McKenna 2012; Shoemaker 2015; Nelkin 2016b; Brink 2018, 2021). I will put side these controversies for now, but they will play important roles in an approach to be presented later in §2.4.3.

2.3   The Pluralist Challenge

In this section, I will raise what I will refer to as a "pluralist challenge" against the communication argument. The central claim is that the communication argument, when combined with the plausible pluralist view that blame has more than one aim, seems to lead to some unwanted consequences. The challenge for friends of the communication argument, then, is to identify a strategy that can accommodate the communication argument with the following thesis:

Aim Pluralism. Blame has more than one aim.

To the extent that defenders of the argument *assume* or *imply* that communication is the only aim of blame, one might think that aim pluralism outright contradicts the communication

---

[4]Other proponents of context-specific responsible agency include defenders of responsibility variantism (Doris, Knobe, & Woolfolk 2007; Knobe & Doris 2010; also see Nichols 2015). For example, Knobe and Doris (2010) cite intriguing empirical evidence in support of treating responsible agency differently depending on whether an individual is a stranger or one we stand in close relationships with.

argument. But this reasoning is too fast. Although defenders of the communication argument rarely acknowledge aim pluralism, it is possible that a charitable reinterpretation of their argument could easily accommodate the thesis. My goal in this section is to show that there is a real challenge here: (i) we have good reasons to believe that aim pluralism is true, and (ii) there is a serious tension between the communication argument and aim pluralism. To do so, I will identify two different but related theses underlying aim pluralism: *concept pluralism*, according to which there is more than one sense, or more than one *concept*, in which blame has an aim (§2.3.1); and *content pluralism*, according to which blame has multiple aims even when we settle on a particular concept of aim (§2.3.2).

2.3.1 Concept Pluralism

One way to establish aim pluralism is to argue that more than one *concept* can be involved in the use of aim-talk:

Concept Pluralism. There is more than one sense in which blame has an "aim."

My proposal is to defend concept pluralism by distinguishing between three concepts of blame's aims: its intentional aim, functional aim, and normative aim. This distinction draws on a close analogy from the literature about the aim of belief. It is widely agreed that beliefs aim at truth, but philosophers have carefully distinguished between different concepts of "aim" that can be involved in such a claim.[5] I suggest that a similar taxonomy can be applied to the case of blame as well, in a way that will turn out to illuminate our discussion of the communication argument.

Theorists have offered three competing kinds of interpretations of the platitude that belief aims at truth: the intentional interpretation, the functional interpretation, and the normative interpretation. The intentional interpretation understands the claim as stating that an agent only

---

[5]See, e.g., Velleman (2000), Wedgwood (2002), Owens (2003), Shah (2003), Shah & Velleman (2005), Steglich-Peterson (2006), Fassio (2011), and McHugh (2012).

believes that *p* when she accepts *p* with the intention of obtaining truth. It follows that believers always have an intentional goal: to accept a proposition only when it is true. The functional interpretation understands the claim as stating that obtaining true beliefs, and avoiding false ones, is the function of the psychological mechanism underlying our beliefs. This function may, but need not, rise to the level of the agent's intention. Both the intentional and the functional interpretations are sometimes dubbed "teleological" interpretations.[6] They are attempts to offer naturalistic, non-normative explications of the claim that belief aims at truth. By contrast, there is the normative interpretation that understands the claim as stating an epistemic norm that is "essential and constitutive for belief" (Fassio 2011, 471): a belief is correct or appropriate if and only if its content is true. According to this interpretation, the expression of "aiming at truth" is best "interpreted as a metaphor" (Wedgwood 2002, 267), meant to capture this epistemic norm.[7] Disagreements between teleological and normative interpretations persist in the literature.

My proposal is that we can distinguish between different senses of the "aim of blame" in a similar way, which provides different ways to explicate and reconstruct the communication argument. First, *X* is an *intentional aim* of blame if and only if an agent always, or at least typically, has an intention to *X* when she blames. The claim that blame aims at communicating a moral demand, then, implies that we blame with an intention to communicate such a demand with the wrongdoer. Second, *X* is a *functional aim* of blame if and only if a function of blame as a type of practice is to *X*. The function of blame may, but need not, rise to the level of intention of the blamer. Instead, the claim that blame aims at a distinctive kind of moral communication implies that it is a function of blame to facilitate this kind of communicative processes. Lastly, *X* is a *normative aim* of blame if and only if blame as a type of practice is governed by a norm like the following: blame is "appropriate" only if it is directed at agents who can *X*. This last interpretation takes the aim of blame as a metaphor. The claim that blame aims at communicating

---

[6]For teleological interpretations see, e.g., Velleman (2000), Steglich-Petersen (2006), and McHugh (2012). See Owens (2003), Shah (2003), and Shah and Velleman (2005) for important objections.

[7]For the normative interpretation see, e.g., Wedgwood (2002), Shah (2003), and Shah & Velleman (2005).

demands, then, implies that blame is appropriate only if it is directed at agents who can fulfill the required role in a distinctive kind of moral communication.

Given this taxonomy, we have good reasons to accept concept pluralism. More accurately, there can be an intentional version, a functional version, and a normative version of the claim that blame "aims" at a distinctive kind of moral communication. How does this challenge the communication argument? Although defenders of the communication argument rarely acknowledge the ambiguity of aim-talk, it is possible that a charitable reinterpretation of their argument would clearly support one reading over the others. Therefore, concept pluralism does not directly pose any objection to the communication argument, but it does invite theorists to take special care to convey the meaning that they intend when contending that *X* is an aim of blame. Note that some instances of aim-talk by defenders of the communication argument—for instance, the claim that blame would be "pointless"—seem open to all three kinds of interpretations. One important exception is Macnamara (2015a, 2015b), who clearly identifies the functional aim of blame as the concept relevant to the communication argument; but the potential ambiguity remains an issue for at least some defenders of the argument. Take Darwall (2006) as an example. On one hand, he frequently puts it that communicating the second-personal demand is what blame "seeks." He writes, for example, that reactive attitudes "seek respect" and they "seek to engage the other second-personally, and that they succeed when the other takes up the address, acknowledges its terms, and thereby respects the dignity of the addresser, both the demand she addresses and her standing to address it" (Darwall 2006, 86). The most natural reading here seems to be intentional or functional. However, Darwall also frequently says that communication figures in the "normative felicity condition" of appropriate blame. As he puts it, a "normative felicity condition" of second-personal address is "a necessary condition of the second-personal reasons actually existing and being given through address" (Darwall 2006, 56). This is more like saying that communicating the demand is the normative aim of blame. Note that the point here is not that Darwall's account is inconsistent; I am inclined to think that his argument can be reconstructed

in a way that avoids any potential linguistic ambiguity. Instead, the point here is that attending to concept pluralism can help clarify and reinterpret some existing theories and arguments that involve aim-talk—otherwise, we risk trading between claims about different concepts of blame's aim.

One reason why we should be especially cautious of this risk, I think, is the following: the normative aim of blame should *not* be the relevant concept that figures in the communicative argument, so defenders of the argument should pay special care to dispel this interpretation. Otherwise, the argument will be the following:

> **The Communication Argument (normative)**. Communicative capacities are necessary for morally responsible agency because a normative aim of blame consists in a distinctive kind of moral communication.

There are a couple of reasons why this should not be how communicative theorists reconstruct their argument. First, I am inclined to think that the normative version of the argument is parasitic on the functional version of the argument: communication is a normative aim of blame *because* it is a functional aim of blame. As I will try to argue later (§2.4.3, §3.4.1, §4.5), blame's functional aims have significant normative implications when it comes to whether we ought to blame. However, we must clearly distinguish the *concept* of a functional aim from that of a normative aim, even if the former aim ends up having certain normative implications. What if one attempts to defend the normative version of the communication argument independent from the functional and the intentional versions? My worry then is that the argument may not be very informative and may lack dialectical value. Its crucial premise is that communication is a normative aim of blame, but that just means that there is such a norm that blame is only appropriate if it targets individuals who possess communicative capacities. Those who do not yet find the conclusion of the communication argument convincing are unlikely to simply accept this premise; the explanatory circle between the premise and the conclusion is too small to make the argument very informative.

Second, the normative version of the communication argument would potentially undermine a Strawson-style response-dependence project. This project is at least important for Watson (2004), who attempts to find a justification for the exemption condition of moral responsibility that is *internal* to our blaming practices involving reactive attitudes. Watson's goal there is to develop and critically evaluate the best possible version of what he calls Strawson's "expressive theory of responsibility" (Watson 2004, 226). The basic tenet of the expressive theory is that "it is not that we hold people responsible because they *are* responsible; rather, the idea (*our* idea) that we are responsible is to be understood by the practice... of expressing our concerns and demands about our treatment of another" (Watson 2004, 222). But if the starting point of the communication argument is a normative claim about when blame is appropriate, then we may worry that the account does not really offer an *internal* justification—the normative claim requires further justification, which may well turn out to be external to our practices of holding people responsible.[8]

Again, concept pluralism does not directly pose any objection to the communication argument, but it invites conceptual clarity. More specifically, the upshot is that theorists defending the communication argument should (i) adopt the intentional version of the argument, (ii) adopt the functional version of the argument, or (iii) appeal to a sense of aim other than that of the intentional, functional, or normative aim. (iii) is a possibility worth more investigation in the future. For now, I will focus on the intentional and the functional versions of the argument. I will argue that, even if we settle on which concept of aim is involved in the communication argument, there is still a further difficulty: blame plausibly has more than one intentional aim and more than one functional aim, which comes into tension with the communication argument.

---

[8]One way to offer a set of criteria internal to our blaming practices about whether blame is "appropriate" is to adopt some response-dependence or fitting attitude account (e.g., Shoemaker 2017), but to appeal to an account of this sort is to locate the source of support for communicative theories elsewhere, instead of in the communication argument itself.

2.3.2 Content Pluralism

Start from the intentional version of the argument:

The Communication Argument (intentional). Communicative capacities are necessary for morally responsible agency because agents who blame always, or at least typically, have an intention to engage in a distinctive kind of moral communication.

This intentional formulation, I think, remains a reasonable reinterpretation of the communication argument to the extent that defenders of the argument take the communicative aim of blame as stemming from the nature of blame as a *demand*. After all, a literal demand qua speech act typically accompanies an intention to achieve some goal. More generally, participants of speech act activities typically intend to participate in such activities. To the extent that blame is thought of as a demand analogous to a form of speech act, it is reasonable to think that the aim of this activity is determined by the intention of the blamer. So one interpretation for the communication argument as defended by Watson, Darwall, and Shoemaker can be as follows: blamers intend to communicate a moral demand, and, therefore, only those who possess the (cognitive, behavioral, and/or affective) capacity to engage in this communicative process are morally responsible agents.

One difficulty with the intentional version of the argument concerns cases of private blame. It seems that many of those who have hidden resentment or indignation do not intend to communicate their demand; their intention is typically to not communicate it.[9] But there may be available moves to deal with this difficulty. One possible move can be found again in the literature about the aim of belief. In his defense of the intentional interpretation, Steglich-Petersen (2006) suggests that there can be "subintentional surrogates" (510) of the intention to obtain truth, if a mental state shares certain characteristics such that it can be conducive to the "hypothetical aim" (515) of someone intending to obtain truth. Similarly, a defender of the communication argument may turn to rely on the weaker assumption that blamers either have an actual intention

---

[9]For a similar argument, see Macnamara (2015a, 2015b) and Driver (2016).

to communicate demands, or have an attitude or behavior that shares certain characteristics conducive to the hypothetical aim of someone intending to communicate demands.

No matter whether such a move is ultimately defensible, my contention is that the intentional version of the argument faces another serious difficulty: the intention to communicate a moral demand does not appear to be the *only* intentional goal that would count as the intentional aim of the practice. That is, the argument is in tension with the following claim:

Content Pluralism (intentional). Blame has more than one intentional aim.

Remember that *X* is an intentional aim of blame if and only if an agent always, or at least typically, has an intention to *X* when she blames. Let's assume that blamers do typically have an intention to communicate demands. Still, we blame with other kinds of intentions as well, sometimes having little to do with addressing a moral demand of reasonable regard. Depending on the context, some might blame with an intention to stand up for themselves and to promote one's self-respect; some might blame with an intention to signal that they share some values widely accepted in their social community, and to maintain social cohesion; some might blame with an intention to vent their negative feelings.[10] At least some of these intentions seem to appear in a wide enough range of contexts such that we can say that blamers *typically* possess these intentions, and, therefore, they should count as blame's intentional aims. Accepting content pluralism does not lead to any logical contradiction with the communication argument; but it does lead to a serious tension. This is because defenders of the communication argument tend to *privilege* the communicative aim over the other aims, but it is unclear how they justify doing so if communication is just one of the many intentional aims of blame. The challenge for defenders of the communication argument, then, is to show why the intention to engage in a distinctive sort of moral communication is special in a way that it leads to a constraint about morally responsible agency.

---

[10]See, for example, Reis-Dennis on the kind of blame that involves an intention and a desire to "rectify. . . perceived slights and disruptions to the moral order" (2019, 452).

One may reply on the grounds that the communicative intention has a special conceptual connection to blame. It is conceptually contingent that blamers have intentions to promote their self-respect. However, to the extent that they engage in the practice of blame, there is a conceptual requirement that they also have a communicative intention. Here one may also invoke the distinction between distal and proximal intentions (e.g., Mele 1992, 2009; Pacherie 2008). Roughly speaking, distal intentions are about what to do *in the future*, while proximal intentions are about what to do *now* (Mele 2009, 10). It might be suggested that, even when blamers have the distal intention of promoting their self-respect, their proximal intention in blame must be to engage in a distinctive kind of moral communication, such as communicating moral demands.

In reply, consider a case that I will call *Idiosyncratic Intention*: Matt was told by his therapist that practicing anger could relieve stress and be good for his mental health. The day after, Matt blames his friend for failing to fulfill a promise, with the intention to act on the therapist's suggestion. Let's grant, for the sake of argument, that Matt also has a proximal intention to communicate a moral demand to his friend. But let's also suppose that the former intention, rather than the latter, is Matt's primary intention: he renders it much more significant in his practical reasoning, and he cares about his mental health much more than whether he successfully conveys a moral demand to his friend. My intuition is that, to the extent that blamers' intentions provide a constraint for who counts as the proper targets to blame, what matters in *Idiosyncratic Intention* is Matt's primary, not secondary, intention. However, the intentional version of the communication argument suggests that Matt's secondary intention, rather than his primary intention, determines the standard for whether his friend is a morally responsible agent. But what explains this? The fact that Matt's secondary intention happens to be one that has some conceptual connection to blame does not seem enough. Generally speaking, even if we take for granted that there is such a conceptual connection and it explains why the communicative intention is a special intentional aim, we still lack an adequate explanation for why it is *special in terms of its normative implications about morally responsible agency*. In any event, it seems to me that we can imagine

that Matt does not even have the proximal intention to communicate; instead, it is possible that Matt's intention about what he does at the moment of blaming is precisely to improve his own mental health, instead of communicating a demand. If so, then the distal/proximal distinction does not help defenders of the intentional version of the argument. There does not seem to be a strong conceptual connection between blame and a communicative intention after all.

Now turn to the functional version of the argument:

The Communication Argument (functional). Communicative capacities are necessary for morally responsible agency because blame has a function of facilitating a distinctive kind of moral communication.

Macnamara (2015a, 2015b) has defended an argument like this, by focusing on blame's function of "eliciting sincere acknowledgment of fault from the wrongdoer," typically in the form of eliciting guilt (2015a, 222). I will focus on Macnamara in the remainder of this section, but note that this functional formulation is another reasonable interpretation of the communication argument for theorists like Watson, Darwall, and Shoemaker as well. Given the problems with understanding blame as always involving a literal demand qua speech act, it is reasonable to reinterpret their demand-talk as the claim that blame has the function of communicating demands. I expect that my arguments below would also apply to this reinterpretation.

Macnamara explicitly endorses an etiology theory of function. According to the etiology account (e.g., Wright 1973, 1976), a function of M is F just in case that (i) F is a consequence of M's being there, and (ii) M is there in part because it does F. As an example, the function of the heart is to pump, since pumping is a consequence of the heart's being there, and current tokens of the heart are there in part because past tokens of the heart did the work of pumping. Applying the etiology theory to blame, Macnamara suggests that (i) past tokens of blame did evoke uptake of their representational content in recipients, and (ii) this partly explains why we have current tokens of blame now. To defend (i), Macnamara appeals to empirical evidence about the facial signatures of reactive emotions (2015b, 557–60); to defend (ii), she contends that

reactive attitudes and their emotional uptake have made a significant contribution to the "building and healing of the moral community" (2015b, 560–2).

But there is a complication. We need to distinguish between the claim that evoking emotional uptake is *the only function* of blame, and the claim that it is only *one of the many functions* of blame. Macnamara means the latter. As she puts it,

> To be clear, the question here is not, "Is this the reactive attitudes' only function?"…
> I take it as a given that reactive emotions, like emotions more generally, have any number of intrapersonal functions—informing the emotion-bearer about, and preparing her to respond to, her environment, to name a few. Our question, then, is whether the reactive attitudes also have the interpersonal function of evoking uptake of their representational content in a recipient. (2015a, 219)

And Macnamara is right to focus on the weaker claim, because it is better supported by the evidence she cites. The fact that reactive attitudes have, through evoking emotional uptake, played a significant part in the building and healing of moral community is good evidence that evoking emotional uptake is one important function of blame, but it is fairly weak evidence for its being the only function. Macnamara herself mentions various intrapersonal functions like "informing the emotion-bearer about, and preparing her to respond to, her environment" (2015a, 219). We thus have good reasons to accept content pluralism about blame's functions:

Content Pluralism (functional). Blame has more than one functional aim.

Though Macnamara seems to assume that content pluralism is perfectly consistent with the communication argument, I think there is in fact a serious tension here. Again, there is no logical contradiction between accepting content pluralism about blame's functions and the communication argument. But defenders of the argument tend to privilege the communicative aim of blame. Given this, the challenge is for defenders of the communication argument to explain what is special about the communicative function that it leads to a constraint about morally responsible agency, but other functions do not.[11] Is it special in that the communicative function

---

[11]One may reply by allowing additional constraints about morally responsible agency. I will discuss this kind of reply, as well as its potential difficulties, in §2.4.1.

is the only *interpersonal* function of blame, whereas other functions are *intrapersonal*? I do not think this is tenable. Blame has interpersonal functions in addition to evoking emotional uptakes as well. For example, various theorists (McGeer 2013; Kogelmann & Wallace 2018; Shoemaker & Vargas 2021) have argued that blame has an important *signaling* function. Shoemaker and Vargas defend an account according to which the definitive feature of blame consists in its "hard-to-fake costly signaling function" (2021, 591). A crucial feature of their account is that they take blame to signal "many types of information" and to "multiple targets" (2021, 590). For example, they argue convincingly that blame functions to signal, first, one's commitment to a relevant norm, and, second, one's own competence in detecting norms in a relevant domain. Along a similar vein, McGeer suggests that blaming emotions are "expensive signals" that are "hard to generate under conditions that do not naturally prime a person to experience them" (2013, 182).

Importantly, blame as a signal can be properly addressed to a recipient even if its recipient lacks the capacity to feel guilt. One can signal that one is *committed* to a norm by issuing blame, and a recipient can properly receive this signal as long as she is sensitive to the relevant contextual cues. The recipient does not need to be able to understand that the content of the relevant norm is *correct* in order to be able to get the signal that one is *committed* to that norm. Feeling guilt, by contrast, requires understanding the force of the content of the relevant norm. If we run an argument structurally parallel to the communication argument, it will go as follows: the capacity of receiving signals is necessary for morally responsible agency because blame has the signaling function. It seems that defenders of the communication argument would also need to accept this signaling argument. However, this then seems to undermine the alleged centrality of communicative capacities to morally responsible agency.

At this point, some may want to deny that blame has a distinctive signaling function. But the more general point stands anyway: blame plausibly has other interpersonal functions in addition to its communicative function,[12] and defenders of the communication argument need to

---

[12]Apart from the signaling function, blame's functions may also include a function of directing attention and epistemic efforts, a function of developing and expanding agents' rational capacities, a function of properly modifying

either (i) explain what is special about the communicative function, or (ii) expand the scope of their argument by adding other constraints about responsible agency. Either way, it is not that easy to retain the force of the functional version of the communication argument while at the same time accepting content pluralism about blame's functions.

Another potential response is that signaling is just a form of moral communication or a part of what moral communication consists in. It might be suggested that signaling one's moral status is communicating one's moral status in some sense, and the point of issuing such signals is for the sake of engaging in moral communication. Thus, it might be suggested, what I've characterized as a pluralist functionalist picture is in fact just a redescription of the communicative view, when different forms and elements of the communicative process are spelled out. However, I do not think that signaling can be categorized as a communicative function. The point of communication, at least according to Watson and Macnamara, is to make the blamee *understand* that she has done something morally bad or morally wrong—for Watson, for the blamee to understand the moral demand; and for Macnamara, for the blamee to understand that she ought not to perform the relevant conduct and thereby feel guilt. Thus, the blamee needs to have the cognitive capacity to understand the relevant norms. But, again, signaling does not presume this cognitive capacity. There is no denying that the signaling function often works together with blame's communicative function. However, they can also come apart, and my contention is that it is a non-trivial task to reformulate the communication argument to accommodate this possibility.

Maybe there is a way to appeal to a broad definition of "communication" such that all interpersonal functions of blame fall into that category. For example, one may argue that signaling and communication (in the sense of communicating demands or evoking guilt) are both forms of *communication for the sake of norm enforcement*. But I worry that this still won't capture all functions of our blaming practices. One crucial function of blame involves *protesting* against a culpable wrongdoing. While moral protest often has a role in norm enforcement as well, protest

personal relationships, etc. The pluralist challenge is a real challenge as long as (i) blame has at least one of these functions, and (ii) this function is distinctive from blame's communicative function.

has a distinctive *expressive* or *symbolic* aim that is not merely for the sake of enforcing norms. For example, the kind of protest involved in blame is conducive to the formation of one's practical identity and to the promotion of one's self-respect.[13]

Further, defenders of the communicative view (e.g., Watson 2004, 2011; Shoemaker 2007) usually want to exempt psychopaths from blame, but even psychopaths may count as possessing communicative capacities if we appeal to a fairly broad sense of communication. This would at least be a striking result for anyone sympathetic to the communicative theory of responsible agency. The takeaway from the pluralist challenge is not that the communication argument is definitely refuted, but that a careful reconstruction of the argument is needed, which may require significant modifications to the communicative view of morally responsible agency as standardly conceived.

2.4 Possible Replies to the Pluralist Challenge

In this section, I will discuss three possible replies to the pluralist challenge. I will focus exclusively on the functional version of the argument, which I take to be the most promising variant, and explore how it can potentially accommodate the pluralist claim that blame has multiple functions.

2.4.1 The Conjunctive Strategy

One natural response to the pluralist challenge is simply to accept the implication that the communication argument needs to be expanded in a way that requires us to add further constraints on morally responsible agency:

The Conjunctive Strategy. *S* is a morally responsible agent only if blaming *S* would fit all functions of blame.

---

[13]I will explain the protest function of blame more carefully later in Chapter 3.

I have said that this threatens to undermine the alleged centrality of communicative capacities to morally responsible agency, but one may not regard this as a huge problem. After all, the communication argument, even as originally presented, is supposed to offer only a necessary condition for moral responsibility. What's the problem, one may ask, if we just allow more necessary conditions?

One worry is that the resulting requirement about responsible agency would seem ill-motivated. The conjunctive view requires a responsible agent to be a fitting target when it comes to all functions of blame. However, if blaming an individual already fits one function of blame, why isn't this enough for that individual to be eligible for the role of blamee? Why should it matter whether it fits *another function* of blame? As an analogy, a ring has both a function of indicating marital status and a decorative function. Imagine a context, like a fashion show, where wearing a ring is only fitting for the decorative function but not for the function of indicating marital status. It still seems appropriate for the model to wear a ring; the complaint that "but this does not fit the other function of rings!" is ill-grounded. There are similar cases when it comes to blame. Imagine an employee, Erin, in a non-ideal workplace, where she is routinely asked to take on extra tasks, doing more work than she is paid for. Also imagine that she has an uncommunicative boss, Pam, who believes that "there is no such a thing as too much work." Given this setting, it seems possible that Pam fails to be a fitting target when it comes to blame's communicative function, but she remains a fitting target for blame's signaling function. Even if communication is futile, Erin can still signal to Pam that she is committed to the relevant norms by blaming Pam. Erin's blame would not fit all functions of the practice. However, her blame still seems appropriate, in contrast to what the conjunctive strategy entails. After all, Erin's blame fits the signaling function, so the complaint that "but blaming Pam does not fit that *other* function of blame!" seems beside the point. Note that this is all consistent with thinking that we may have other, non-function-related grounds to think Pam ought not to be blamed, for example, by arguing that she is non-culpable, does not deserve blame, etc. But to the extent that we rely on

the communication argument, or a variant thereof, to draw conclusions about responsible agency, the conjunctive strategy seems to lead to an overly strong and ill-motivated requirement.

2.4.2   The Disjunctive Strategy

If the conjunctive strategy seems too strong, then how about a disjunctive approach? The idea will be as follows:

> **The Disjunctive Strategy**. *S* is a morally responsible agent only if blaming *S* would fit any one of the many functions of blame.

The natural worry here is that the criterion is too weak. In the absence of other constraints for responsible agency, it implies that one is eligible for the role of blamee as long as one has the capacity to, for example, receive signals about norm commitments. But this is a really low bar. Again, even psychopaths—who many communicative theorists (Watson 2004, 2011; Shoemaker 2007) tend to exempt—would count as morally responsible agents. We would also include young children who are developmentally mature enough to receive moral signals, but not yet mature enough to engage in the communication of moral demands or have the capacity to feel guilt. Further, the disjunctive strategy seems to render communication—either in the sense of communicating demands or evoking guilt—only one of the many elements that we need to attend to in deciding who is eligible for the role of blamee. There would seem nothing special at all about communication in this process; but the uniqueness of the communicative function of blame is an important part of why the argument becomes appealing in the first place.

2.4.3   The Context-Specific Strategy

Both the conjunctive strategy and the disjunctive strategy lead to some unwanted consequences. There is more to be said about these two strategies, and perhaps they can be developed in more careful ways to avoid the potential problems. But I think there is another strategy available:

> **The Context-Specific Strategy**. *S* is a morally responsible agent in a distinctive kind of context *C* only if blaming *S* would fit the contextually significant function in this kind of context (when there is only one contextually significant function in *C*).

The general idea is this. When a type of practice has more than one function, and to the extent that its functions constitute restrictions of the propriety of the practice, these restrictions ought to be *context-specific*. Consider the example about ring-wearing again. It is intuitive that whether ring-wearing is appropriate in a fashion show has little to do with whether it would fit the function of signaling marital status.[14] The natural explanation is that the latter function is not significant in this context. Instead, the contextually significant function is the decorative function. Thus, the fittingness of ring-wearing depends only on this contextually significant function; whether it fits some other function of the practice is simply beside the point.

The context-specific strategy is based on a similar thought. The communicative function is only one of the many functions of blame, but there is a distinctive kind of context—let's call this the "communicative context"—where this function is contextually significant. We then may be able to explain why the communicative function is special by explaining the uniqueness of this kind of context.

But there is a further problem. What if there are two or more equally or similarly significant functions in *C*? I am inclined to think that the formulation of the strategy then needs to take into account degrees of responsible agency. If blaming *S* would fit all contextually significant functions, *S* is a *fully* morally responsible agent in *C*; but if blaming *S* would fit some but not all contextually significant functions, then *S* has diminished morally responsible agency in *C*. The account then needs to be modified as follows:

> **The Context-Specific Strategy (scalar)**. *S* is a morally responsible agent in a distinctive kind of context *C* only to the extent that blaming *S* would fit the contextually significant function(s) in *C*.[15]

---

[14]Of course, rings can be used in fashion shows to make claims about desirable fashions for married persons as well. But even in such a context, whether the model who wears a ring is herself married, again, seems to have little to do with whether it is appropriate for her to wear the ring.

[15]One might wonder why we need to appeal to context-specific agency at all, if we already treat responsible

I find the context-specific strategy promising. But what exactly decides the contextually significant function(s) in C? One way to do so is to examine the function(s) that we, either based on some convention or based on some other shared understanding of an expectation, associate with C.[16] For example, the decorative function of wearing a ring is conventionally associated with the context of a fashion show. On this approach, the communicative function is a contextually significant function of blame in C when there is a convention or a shared understanding of an expectation that communication is an important goal in C.

Another way to decide the contextually significant function(s) is to examine the function(s) that should matter in C, given the morally salient features of this context. The thought is that different kinds of contexts can highlight and make salient different elements of one's situation that require our moral attention, which can then make a difference about the contextually significant function(s).[17] I think we should combine those two approaches to understanding contextual significance. Sometimes the communicative function becomes contextually significant due to our associating the function with C, on the basis of convention or shared understanding of an expectation. At other times it becomes contextually significant because communication is an important goal given the morally salient elements of C that require our attention. Both kinds of circumstances constitute the "communicative context"—those circumstances where the communicative function of blame is contextually significant.

We can then have a version of the communication argument according to which the communicative function affects, to a significant degree, the extent to which one is a morally

---

agency as coming into degrees. Why not just modify the conjunctive strategy such that S is a morally responsible agent only to the extent that blaming S would fit blame's functions? I think that this modified conjunctive strategy still faces a serious problem, because there are cases where, as I mentioned in §2.4.1, a function is not contextually significant and, intuitively, should not affect even one's degrees of responsible agency.

[16]For related discussion, see Walker on different senses of a "context," including the sense that "invokes a shared understanding of the expectations that are in play in common social encounters" (2002, xi).

[17]This is related to Herman's (1993) discussion about "rules of moral salience" in moral knowledge. According to Herman, rules of moral salience are what enable a person to "pick out those elements of his circumstances or of his proposed actions that require moral attention" (1993, 77). I am agnostic about whether such rules exist or how they figure in moral epistemology. But I agree with Herman that the morally salient features of a situation should play a crucial role in our moral theory.

responsible agent in the communicative context. We may also preserve the uniqueness of the communicative function by defending the uniqueness of the communicative context. More specifically, the communicative context seems to be fairly *pervasive*: in many, if not most, blaming contexts, communication is at least one of the contextually significant functions. On one hand, we often associate the communicative function with blaming contexts, especially when we blame people whom we love or care about; we typically expect blame to serve a communicative goal and to initiate a conversation between us. On the other hand, there are also many situations that make salient the *need* for moral communication, for example, when blame targets a norm violation that occurs because of miscommunication or lack of shared moral understandings between the blamer and the blamee. In these circumstances, communication should be an important functional goal given the features of the situation that deserve our moral attention. The pervasiveness of the communicative context, I think, can potentially justify why one can privilege the communicative aim of blame over the other aims.

However, the context-specific strategy also allows contexts where the communicative function is not significant, or significant only to a fairly small degree. This can be what happens between Erin and Pam. I take it that the morally salient features of the context are the non-ideal features of the workplace. The need for communication is much less salient, especially given the radical asymmetry between the communicative power of the potential interlocutors. This can explain why Erin's blame still seems appropriate: blame is fitting for its most contextually significant functions, including the function of signaling Erin's commitment to norms. The context-specific strategy thus allows the possibility that one lacks the communicative capacity but is still a partially or even fully morally responsible agent.

The conclusion of the context-specific strategy relies on the assumption that morally responsible agency is context-specific. As I have mentioned in §2.2, there are good reasons to accept a context-specific or "patchy" picture of responsible agency (see, e.g., Vargas 2013), but this remains a controversial matter. My formulation of the strategy also assumes that responsible

agency comes into degrees. More work needs to be done to defend these assumptions, but the context-specific strategy seems a more promising way out of the pluralist challenge compared to its alternatives. Despite its relatively moderate stance on the centrality of communication to responsible agency, it may provide the best reconstruction of the communication argument in a way that preserves much of its theoretical insight.

2.5 Concluding Remarks

I have proposed the pluralist challenge: despite the appeal of the communication argument, it is in tension with the plausible pluralist claim that blame has more than one aim. I have discussed some possible strategies that one may adopt to respond to the challenge and the difficulties facing each, and I have suggested that the context-specific function-based approach seems most promising. The upshot is that accommodating the communication argument in a pluralist picture about blame's aims is no easy task, and to do so may require its proponents to abandon or at least modify the initial stance about how central communicative capacities are to morally responsible agency.

Another issue worth further investigation is how to understand the aims of those practices related to or resembling blame. For example, praise and blame are often treated symmetrically in the literature, but do they have the same aims or not? How to think about the relations between the aims of blame and the aims of forgiveness? How much analogy or disanalogy there is to make between the aims of blame and the aims of criminal punishment? These are all interesting issues, and I expect that a pluralist view about aims similar to the one proposed in the current chapter can help shed light on them.

Chapter 2, in part, has been accepted for publication (as "The Communication Argument and the Pluralist Challenge") in Canadian Journal of Philosophy, 2021, Wang, Shawn Tinghao. The dissertation/thesis author was the primary investigator and author of this paper.

Chapter 3

Blame: A Pluralist Function-Based Approach

3.1   Introduction

What is blame? Three kinds of accounts are among the most popular: content-based accounts, emotion-based accounts, and function-based accounts. According to content-based accounts, blame is defined primarily in terms of a psychological state (or a set of psychological states) and its characteristic representational content (or their representational contents), such as a belief that a person has shown a substandard quality of will (Hieronymi 2004) or a belief-desire pair, including a belief that a person has acted badly and a desire that she not have performed that act (Sher 2006). By contrast, emotion-based accounts define blame primarily in terms of an emotional attitude—more specifically, in terms of what Strawson (1962) calls "reactive attitudes" such as guilt, resentment, and indignation. Proponents of emotion-based accounts typically suggest that the representational content of the emotional attitude is not the most central—or, at least, not the only central—element of blame. Instead, blame is defined by an emotion and its various properties, including its phenomenology, representational contents, tendencies of

interpretation, and associated psychological and behavioral dispositions.[1]

Both content-based accounts and emotion-based accounts, however, appear to many as being too limited to account for the diverse phenomenon of blame. Specifically, many share the intuition that a pre-theoretically appealing account of blame should include both emotional forms of blame and non-emotional, dispassionate forms of blame. Function-based accounts are meant to capture this intuition by defining blame primarily in terms of the functional role that it serves. In this respect, blame is defined "more like a mousetrap than a diamond" (Portmore 2022, 49; see Polger 2021). Since the very same functional role can be realized in different ways, these accounts allow blame to have both emotional and non-emotional manifestations.[2] While a function-based account is thereby deeply appealing, it also raises new difficulties, and disagreements among the content-based, emotion-based, and function-based accounts persist in the literature.[3]

The goal of this chapter is to contribute to this debate by developing a novel *function-based* approach to blame. In particular, my function-based approach is *pluralist*. On my favored version of the approach, blame as a type of practice is defined by a cluster of non-accidentally associated functions, including its protest, communicative, and signaling functions. For *A* to blame *B* on a particular occasion, on this view, is just for *A* to perform an action or hold an attitude toward *B*, where this action or attitude has a robust disposition to serve enough of the functions in this cluster. After motivating the approach by comparing it to existing function-based views (§3.2), I develop my favored version of the approach that involves blame's function-cluster (§3.3). I then respond to two objections (§3.4). I conclude that a pluralist function-based approach has various

---

[1]Defenders of emotion-based accounts include, for example, Wallace (1994, 2011), Wolf (2011), Tognazzini (2013), Pickard (2013), Cogley (2013), Rosen (2015), and Menges (2017a).

[2]Defenders of function-based accounts include, for example, McGeer (2013), Fricker (2016, 2018), Shoemaker and Vargas (2021), and Vargas (2021). Other accounts that can be reasonably interpreted as function-based include, for example, Scanlon (2008), McKenna (2012), and Smith (2013).

[3]This is not an exhaustive list of all possible theories of blame. For example, McKenna (2012) endorses the view that blame consists of primarily the outward blaming behaviors. There are also many theorists who endorse hybrid views that have, for example, both cognitive and functional elements (e.g., Brink & Nelkin 2022) or both cognitive and emotional elements (e.g., Portmore 2022).

advantages and should be a serious contender in the debate about the nature of blame.

3.2   Motivations for a Pluralist Function-Based Approach

A pluralist function-based approach to blame is a theory that accepts the following three claims:

(I) **The functionalist thesis**. Blame is defined primarily in terms of its functions.

(II) **The pluralist thesis**. Blame, as a type of practice, has more than one function.

(III) **The non-derivative thesis**. More than one function of blame plays a non-trivial and non-derivative role in the analysis of blame.

By "non-derivative," I mean that the role of blame's one function *F1* is not simply derived from the role of another of blame's functions *F2*. A theory does not count as being pluralist function-based if it allows blame to have multiple functions but simply derives all their roles in the definition of blame from just one of these functions.

The pluralist function-based approach is similar to existing function-based accounts in various ways, but it is prefaced by the pluralist thesis: blame, as a type of practice, has more than one function. Though various theorists have mentioned or briefly discussed the pluralist thesis, there has not been any systematic examination of how it can be integrated into a function-based approach to blame. Rather, proponents of function-based accounts tend to assume or write as if blame has one single, unifying function, despite their disagreement on what this unifying function is. Shoemaker and Vargas, for example, argue that this is a function that involves "signaling of the blamer's commitments, including a commitment to the enforcement of those commitments" (2021, 582), and they refer to this signaling function as the "unifying core feature of blame" (2021, 591). McGeer suggests that blame's function involves action-priming, consisting in its tendency to be triggered by a "transgressive" behavior and to induce an "aggressive (punitive) response aimed at changing or inhibiting such behavior" (2013, 172). And Fricker contends that the "basic point and purpose" of blame consist in its communicative function (2016, 171). Blame

initiates and facilitates communication about our shared reasons among members in the moral community, and it works as "an invaluable interpersonal calibrator in moral agency" (Fricker 2016, 177). These theorists formulate function-based accounts of blame that can potentially allow multiple function realizers, but they nonetheless often assume a monist stance about the functional role that is constitutive of blame.

However, we should at least be open to possibility that blame turns out to have more than one function. Let me explain why. Theories of function can be divided into two kinds: historical theories and ahistorical theories. Historical theories purport to understand the function of an entity-type in terms of what has happened in the past, such as the history of this entity-type's selection, reproduction, reinforcement, or persistence.[4] For example, according to Wright's (1973) etiological theory, a function of E is F if and only if current tokens of E exist in part because enough of past tokens of E did F. A function of zebra stripes is to deter flies, then, if and only if current tokens of zebra stripes exist in part because enough of past zebra stripes did the work of deterring flies. As a more recent example of a historical theory, Garson has defended a "generalized selected effect theory" of biological functions: a function of a trait is "an activity that led to its differential reproduction, or its differential retention, in a population" (2019, 93). Both etiological and selected effected theories allow an entity-type to possess more than one function. For instance, Garson notes that, according to selected effect theories, a trait can have multiple functions if "it's involved in multiple selection processes" (2019, 100). The male urethra has both the function of transporting semen and the function of transporting urine (Garson 2019, 118); the brain's medulla has both the function of contributing to breathing and the function of contributing to the gag reflex (Garson 2016, 70). How about blame? Macnamara makes the point that blame has more than one function according to an etiological theory, including both a communicative function and "any number of intrapersonal functions" such as "informing the emotion-bearer about, and preparing her to respond to, her environment" (2015a, 219). More

---

[4]For defenders of historical theories of functions, see, e.g., Wright (1973), Millikan (1989), Neander (1991a, 1991b), Buller (1998), McLaughlin (2001), Bouchard (2013), Garson (2019).

generally, the pluralist thesis is appealing if historical theories of functions are applicable to blame. After all, if a function of blame as a type of practice is to be understood in terms of what can in part explain the selection, reproduction, reinforcement, or persistence of blame, then it is quite plausible that multiple factors contribute to this explanation. For example, the fact that enough of past tokens of blame signaled people's commitment to norms appears to contribute to why blame, as a type of practice, persisted and has been reinforced in its recent history; but so does the fact that enough of past tokens of blame operated as valid protests, and so does the fact that enough of them initiated and facilitated communication among members in the moral community. The natural consequence then is that signaling, protest, and communication should all count as blame's functions. At the least, the pluralist thesis would appear to be a substantial possibility about the history of blame that those sympathetic to function-based accounts should not simply close off.[5]

The pluralist thesis is even more appealing, I think, if it turns out that some ahistorical account of function is applicable to blame. Ahistorical theories of function are based on the intuitive thought that the function(s) of an entity-type is ultimately about what it can potentially do now, regardless of its past history.[6] One prominent example is Cummins's (1975) causal role theory of function: a function of an entity-type is understood in terms of its causal role in the appropriate and adequate account of the proper operating of a certain system. Thus, the heart functions as a pump because the heart plays the causal role of pumping in the appropriate and adequate account of the circulatory system's capacity to transport nutrients and oxygen. The causal role theory of function naturally allows an entity-type to possess multiple functions, since a given entity-type often "plays a role in many different systems, and hence it has a large number of functions" (Garson 2016, 83). Blame in particular appears to play multiple roles in the system

---

[5]One may respond that signaling, protest, and communication all concern *communication for the sake of norm enforcement*, so the argument here in fact supports monism rather than pluralism. I will address this reply later in §3.4.2.

[6]For defenders of ahistorical theories of functions, see, e.g., Cummins (1975), Boorse (1976), Bigelow & Pargetter (1987), Craver (2001), Hardcastle (2002), Nanay (2010).

that involves moral practices in general. For example, blame promotes norm enforcement through its role of signaling individuals' commitments to norms, honors moral values and individual rights through its role of issuing protests, and promotes mutual understanding through its role of initiating and facilitating communication among us. The natural consequence is that signaling, protest, and communication should all count as functions of blame. This, again, lends support to the pluralist thesis.

Another, more dialectical, motivation for pursuing a pluralist function-based approach is as follows. A common motivation for function-based accounts is that they allegedly improve upon content-based and emotion-based accounts by allowing the same function of blame to be realized in various forms. Whereas a content-based account states that blame is always a cognitive judgment or always a belief-desire pair, and an emotion-based account states that blame is always a negative emotion, a function-based account can allow that blame has any number of these forms, as long as the relevant psychological states realize the same function. The thought is that functionalists can be pluralists about the *realizers* of blame's function, and in doing so they can better characterize the diverse phenomenon of blame than their opponents. As Shoemaker and Vargas put it, a function-based account is able to avoid worries about extensional adequacy and does not have to "employ ungainly fixes to deflect one or more apparent counterexamples" (2021, 581–2).[7] But if being a pluralist about function-realizers is a major virtue of the function-based approach, then the natural next step is to be a pluralist not just about function-realizers, but also about blame's *functional roles*. This is what the pluralist function-based approach is aiming for. If such an approach is defensible, then it may allow us to accommodate an even broader set of attitudes and activities as blame, and it may do an even better job at characterizing this apparently diverse phenomenon. Being a pluralist invites new worries and difficulties, but the approach is definitely worth trying out, given the potential theoretical payoffs that it may provide.

---

[7]Fricker motivates her function-based account in a similar way, starting from the claim that "the practice of blame is significantly disunified" (2016, 166). Also see McGeer's (2013) argument that content-based accounts often involve dubious attempts to "sanitize" the phenomenon.

Indeed, some function-based theorists have occasionally acknowledged that blame has multiple functions or at least multiple functional dimensions. McGeer, for example, mentions that blame has "an important signaling function" in addition to its action-priming function (2013, 172). Smith states that blame has two functions: "First, to *register* the fact that the person wronged did not deserve such treatment by *challenging* the moral claim implicit in the wrongdoer's action; second, to prompt moral recognition and acknowledgment of this fact on the part of the wrongdoer and/or others in the moral community" (2013, 43).[8] However, it is not entirely clear if McGeer and Smith would also accept the non-derivative thesis. For example, Smith is quick to clarify that it is the first, protest function of blame that is more primary (2013, 43); on one reading of Smith, the second function of blame is merely derived from the first, more primary aim. But these comments on the multi-dimensional feature of blame's function, I think, should at least motivate us to take a pluralist function-based approach seriously. In fact, I suspect that many function-based theorists may, if pushed to answer the question, accept the pluralist thesis and even the non-derivative thesis. Their accounts may thus turn out to be pluralist after all. In any event, there has not been any explicit, systematic examination of the pluralist function-based approach. This chapter can be seen as an attempt to fill in this lacuna. Even if the pluralist approach that I endorse is already nascent in existing views, making the pluralist nature of an account of blame explicit will be theoretically illuminating by highlighting both its potential advantages and distinct difficulties.

3.3   The Cluster of Function Strategy

The goal of this section is to articulate and elaborate a concrete version of the pluralist function-based approach to blame. My proposed view—the "cluster of function strategy"—

---

[8]Also see Cogley (2013) and Macnamara (2015a, 2015b) on the view that a reactive emotion can have multiple functions. Both explicitly endorse the view that blame has multiple functions; however, their focus is on blaming emotions in particular, and they do not attempt to define blame in terms of its functions.

involves an attempt to define blame with reference to a function-cluster.[9] It includes both a definition of blame *as a type of practice* and a definition of blame *as a token attitude or token activity*:

(i) **Blame-type functionalism**. A type of practice T is blame if and only if T's functions constitute a function-cluster C that consists of at least three functions: protest, communication, and signaling.

(ii) **Blame-token functionalism**. A token attitude or token activity K is blame if and only if K has a robust disposition to serve enough of the functions in the function-cluster C.

Both (i) and (ii) need further clarifications, which I will turn to now.

3.3.1 Blame-Type Functionalism

Let's start with (i), the definition of blame-type: it is simply the type of practice that has a cluster of functions including at least protest, communication, and signaling. Most accounts of functions described earlier, historical and ahistorical alike, are meant to be in the first place about functions of types of entities. (i) is largely neutral about which of these accounts is applicable to blame. As I have mentioned in §3.2, it is quite plausible that protest, communication, and signaling all count as functions of blame. They all in part contribute to the persistence, retention, and reinforcement of blame as a type of practice in its recent history, and they are all ways in which blame plays a role in contributing to the broader system involving our moral practices in general.

What exactly are the protest, communication, and signaling functions? Let me explain them in order.

---

[9]My account is inspired by Brink and Nelkin's (2022) "core and syndrome" account of blame. On their view, blame involves both a cognitive core, in the form of a belief that someone is blameworthy, and a non-accidental syndrome, in the form of various sorts of private and public dispositions. They mention the possibility of interpreting syndrome as a "non-accidental cluster of elements" (Brink & Nelkin 2022, 185). Unlike Brink and Nelkin, however, my account does not have a cognitive element.

(a) *Protest*. One function of blame is to issue a protest. Blame protests by, first, registering a truth that someone is culpable of a wrongful conduct, and, then, marking out and challenging either a false claim or a substandard treatment that is implicated in that conduct.[10] Sometimes we protest because people's actions assume or imply that we are "not deserving of moral respect" (Smith 2013, 42), for example, when we blame for racist behaviors. At other times, we protest against a substandard treatment even when the treatment does not appear to assume anyone's lack of entitlement to respect, for example, when we blame for a friend's lack of interpretive generosity about something that we have said. Either way, through protesting, we draw attention to something that *matters* to us, even if the target of our blame does not see the value of what we protest about. As Talbert puts it, the protester "puts something important on record, but it need not be his goal that his oppressors also affirm his moral standing" (2012, 106). The protest function of blame thus involves a primarily *unilateral* form of expression, one that "is meant largely for the protestor and for his fellow sufferers" (Talbert 2012, 106).

(b) *Communication*. Another function of blame is to communicate. Blame communicates by initiating and facilitating a process of moral conversation between the interlocutors, while seeking certain changes of behaviors and/or attitudes in the wrongdoer through this conversation.[11] Consider an example from Wolf (2011, 334). A mother blames her daughter for repeatedly borrowing her mother's clothes without permission. Sometimes blame initiates a moral conversation by itself being part of that conversation, for example, if the daughter talks back, followed by "the slammed doors and raised voices" that are "accompanied by the discussion of these events" (Wolf 2011, 334). At other times, blame triggers a conversation without itself being part of it, for example, if the mother expresses her resentment but says "we will talk about this later." Either way, blame's communication

---

[10]For related discussion on blame and protest, see, e.g., Boxill (1976), Talbert (2012), and Smith (2013).

[11]For related discussion on blame and conversation/communication, see, e.g., McKenna (2012), Smith (2013), McGeer (2012, 2013, 2019), and McGeer and Funk (2017).

is a *directed* process that involves a *mutual exchange of reasons* regarding the moral quality of a conduct. It must be directed in the sense that it is a process in which one communicates with some particular target (an individual or a group). It must involve a mutual exchange of reasons in the sense that the communicative process cannot be simply one-sided; the interlocutors should be on relatively equal footing. This does not mean that they must have the same social status, as can been seen in the example of the mother and the daughter. But the communication must be an *interactional* process. The mother will talk *with*, not just talk *to*, her daughter. Another important element of the communicative function involves seeking certain behavioral and/or attitudinal changes in the target of blame. The mother expects her daughter to stop borrowing clothes without permission, and to be a more responsible adult, in the future. But I do not think that blame's communicative function always seeks a transformation in people's future behaviors. Sometimes the conversation is only meant to bring about a change in people's attitudes, for example, to make people understand that they have done something wrong and feel the appropriate degree of guilt or remorse.[12]

(c) *Signaling*. The third function of blame involves signaling. The idea is that blame is a meaningful social signal that carries information around, such that people who witness a blaming activity or are aware of a blaming attitude could make some inferences about the blamer—about her commitments, capacities, or qualities of will.[13] Blame is a "signaling delivery device" (Shoemaker & Vargas 2021, 589). For example, imagine that Ben's friend tells a joke that makes fun of Asian-Americans, accompanied by the remark "but of course I'm not a racist!" Ben still blames his friend, and in doing so he signals that he is committed to norms against Asian-American racism, that he is capable of recognizing and reacting

---

[12] One possible worry is that this usage of "communicative function" is too narrow, and a broad notion of moral communication can in fact incorporate both protest and signaling as parts of the communicative process. I will address this worry later in §3.4.2.

[13] For related discussion on blame and signaling, see, e.g., McGeer (2013), Kogelmann & Wallace (2018), and Shoemaker & Vargas (2021).

to the relevant, sometimes subtle, norm violations, and that he has good quality of will toward Asian-Americans. People who are aware of the fact that Ben blames his friend, then, can make reasonable inferences about his commitments, capacities, and qualities of will. There is an important difference between signaling and the other functions of blame. While protest is largely meant for the blamer herself, and communication is largely meant for the blamer and the blamee, blame's signal is meant for a broader social community. Ben's friend can pick up the signals, but so can other people in Ben's society.[14]

The protest, communicative, and signaling functions constitute blame's "function-cluster." This is a *cluster* rather than a mere *collection* of different functions, in the sense that these functions are organically related to each other. More specifically, they are mutually reinforcing in a way similar to Boyd's influential notion of a "homeostatic property-cluster" (1988, 1999a, 1999b, 2003a, 2003b). According to Boyd, some entities—natural kinds, properties, and relations—are defined, not by any necessary and sufficient conditions, but by a " 'homeostatically' sustained clustering of properties or relations" (1999a, 141). Boyd refers to them as "homeostatic property cluster kinds" (1999a, 141). A homeostatic property-cluster F consists of a set of properties that satisfy a series of conditions, and one of them is that they are "stable and mutually reinforcing" (Kumar 2015, 2889). This means that, under appropriate conditions, some properties in F are inclined to "favor the presence of the others" or there are "underlying mechanisms or processes which tend to maintain the presence of the properties in F" (Boyd 1988, 323). Another feature of a homeostatic property-cluster is that the properties in the cluster are jointly valuable. This means that, according to Boyd, the joint occurrence of "(many of) the properties in F together with (some or all of) the underlying mechanisms in question" can produce "(theoretically or practically) important effects" (1988, 323).

---

[14]The quality of blame's signal varies. It can sometimes be quite costly. For example, when an employee blames her employer, she does so at the cost of potentially losing her job. But at other times blame's signal is quite cheap. For example, one costs little to nothing when one blames a serial killer whose action everyone agrees is gregariously wrong.

My proposal is that Boyd's notion of property-clusters can be generalized or extended in a way that makes sense of *function-clusters*. Like property-clusters, function-clusters are sets of mutually reinforcing and jointly valuable functions. Blame, in particular, is defined by a function-cluster like this. First, blame's functions *mutually reinforce* each other. The realization of a function in the cluster is in general causally conducive to the realization of others. For example, consider the relation between blame's protest and communicative functions. Moral protest against a wrongful conduct often makes it more likely that we can successfully engage in a communicative process with the wrongdoer, since they are more likely to direct their attention to the protester. Conversely, communication in the form of a mutual exchange of reasons often can help fix the interlocutors' common ground in a way that leads to more considered, targeted moral protest. This relation between protest and communication is similar to what, for instance, a union does when bargaining with the employer about raising workers' salaries—a process that often involves both employee strikes (as a form of protest) and exchanging reasons at the negotiation table (as a form of communication). Protest and communication also reinforce the signal that one is committed to a norm, since the fact that one takes the time and effort to protest against the norm violation and to communicate about the relevant moral reasons strengthens the implication about one's norm commitment. Conversely, when blame functions as a meaningful signal that carries the information that the blamer is committed to a norm, it makes it more likely for the blamer's protest and communicative efforts to prevail. Blame's functions are thus mutually reinforcing: the fulfillment of some functions in the cluster makes it more likely that the other functions would also be fulfilled.

Second, blame's functions are also *jointly valuable*. Boyd suggests that the joint occurrence of "(many of) the properties" in a property-cluster can produce "(theoretically or practically) important effects" (1988, 323). Similarly, I think that the joint realization of (many of) the functions in a function-cluster should be theoretically or practically valuable. What does this consist of? My proposal is as follows: the joint realization of blame's functions can *contribute to*

54

the realization of the aims of the system of first-order normative practices in an efficient and non-alienating way.

Here is what I mean. Blame is, in a certain sense, a second-order normative activity: it is typically a (normative) response to a norm violation. By contrast, there are first-order normative practices such as the norm-following acts and the norm-violating acts themselves. Aside from acts, there may also be omissions, attitudes, and even character traits that are governed by norms. These norm-following and norm-violating practices together constitute a central part of what I will refer to as the *system of first-order normative practices*. These practices often result from people's exercising or failing to exercise their norm-following capacities in particular circumstances. The exact scope of this system is debatable. It is quite uncontroversial to think that it involves practices surrounding concepts such as moral obligations, moral duties, and what is morally right and morally wrong. But this may not be all that the system is about. For example, we may think that blame is not just about moral conduct; we blame not just for morally wrong conduct but also for, for instance, epistemically wrong conduct. We may also think that the scope of this system includes suberogatory actions, attitudes, and character traits; we blame not just for wrong conduct but also for bad conduct. Both matters are, however, fairly controversial, and I will be neutral on them for the purposes of this chapter.

Further, this system of first-order normative practices has various internal aims, including, for example, (a) promoting right rather than wrong actions, (b) honoring moral values and respecting individual rights, and (c) improving agents' norm-following capacities.[15] Given this, my suggestion is that the joint value of blame's functions consists in their positive contribution to these first-order normative aims. And not just any sort of contribution—they contribute to the realization of these aims in an *efficient* and *non-alienating* way. On one hand, blame is quite an efficient kind of practice. When blame realizes its protest, communicative, and signaling

---

[15]Theorists who endorse different first-order normative ethical views (consequentialism, deontology, etc.) will put different emphasis on these first-order aims. My central suggestion is that, no matter what the first-order aims are, blame contributes to them in an efficient and non-alienating way.

functions, it does an extraordinarily efficient job in helping promoting right actions and improving agents' norm-following capacities. In cases like this, blame at the same time marks out and challenges a wrong conduct, initiates a communicative process about the moral quality of the same conduct, and signals one's commitment to the norm that this very conduct violates. A multi-dimensional mechanism is embedded within a single attitude or activity of blame. But blame's contribution is also non-alienating. It does not obtain the first-order aims by coercing, punishing, or manipulating agents. Instead, blame helps realize these aims by connecting them to both the blamer and the blamee in a fairly personal way. The personal connection with the blamer is particularly salient in the protest function, whereas the personal connection with the blamee is particularly salient in the communicative function. When protesting against a norm or being called to exchange reasons about a norm, an individual is more likely to feel *personally connected* to the goals of blame. She is more likely to be invested in those goals and more likely to treat these goals as important in her own individual experiences. By incorporating protest and communication into blame's function-cluster, we can see how blame can make its values feel personal to all parties.

To sum up: Blame, as a type of practice, is defined by its function-cluster. This cluster involves protest, communicative, and signaling functions. These functions are mutually reinforcing and jointly valuable, providing a particularly efficient and also non-alienating pathway into realizing the aims of the system of first-order normative practices.

3.3.2   Blame-Token Functionalism

I have offered a definition of blame-type. But a theory of blame should also help us know whether a given *token attitude* or *token activity* counts as blame. This may appear an easy task. One may think that a token attitude/activity counts as blame simply when it belongs to the blame-type that we just identified. However, this assumes that there is an independent method of type individuation. To illustrate, note that blame does not always realize its functions; an

56

individual token of blame can dysfunction like a broken clock. A problem then occurs: how do we distinguish between (a) a dysfunctional token that still belongs to the blame-type (analogous to the broken clock) and (b) a token that does not belong to the blame-type (analogous to something that is not a clock at all)? More specifically, consider some hard cases for theories of blame, such as blaming the dead. Blaming the dead probably does not realize blame's communicative function. But is this an example of dysfunction, such that blaming the dead still falls under the broader category of blame, or is this an example where the practice does not count as blame at all? It is surprisingly hard for the functionalist to answer this question—she cannot simply respond that there is a natural way of individuating types, since one goal of a theory of blame is exactly to find out how to do this.

My view is that a function-based account can provide a more direct definition of blame-token in terms of *the dispositional properties of this very token*. This is the second part of my definition of blame:

(ii) **Blame-token functionalism**. A token attitude or token activity K is blame if and only if K has a robust disposition to serve enough of the functions in the function-cluster C.

My account is partly inspired by Nanay's modal theory of functions, which involves an attempt to define the function of an entity-token in terms of its behavior in the "relatively close" possible worlds (2010, 422). (ii) similarly appeals to a modal fact about the token. An account would be a non-starter if it defines blame according to whether it actually fulfills a certain number of functions. It is simply not the case that every token of blame actually fulfills protest, communicative, or signaling functions. Blame can dysfunction. An instance of blame can dysfunction by fulfilling some functions in the cluster C while failing to fulfill others. But it can also dysfunction by failing to fulfill all functions in the cluster. As an example, imagine a case of resentment that is *intentionally well hidden* from everyone else. It is so well hidden that it does not issue a protest, it does not initiate any communicative process, and it does not constitute a meaningful social signal. This resentment-token, however, can still intuitively count as blame.

57

My account accommodates this intuition by pointing out that this resentment-token can still have a *disposition* to fulfill these functions. It therefore can still count as a case of blame.

One way to flesh out the notion of a disposition is to appeal to the behaviors of a token in nearby possible worlds. For example, Manley and Wasserman suggest that a thing has the disposition to *M* when *C* just in case that it would *M* in "in some suitable proportion of *C*-cases" (2008, 76). On this view, (ii) then implies that a blame-token would need to fulfill the relevant functions in some suitable proportion of nearby possible worlds. In addition, note that dispositions are usually indexed to a particular context. A token activity may happen to have a disposition to fulfill the relevant functions only in some far-fetched or peculiar context, whereas an activity, as I will use the term, has a "robust disposition" to fulfill a function when its relevant disposition covers a wide enough range of contexts. (ii) requires a blame-token to have a robust disposition. This means that its disposition to fulfill the relevant functions needs to cover a sufficiently wide range of contexts. More generally, the intuition is that a blame-token needs to be *reasonably well suited for* or *good enough at* fulfilling the relevant functions; even when it fails to actually fulfill them, a blame-token has *what it takes* to do so.[16]

(ii), however, does not require a blame-token to be reasonably well suited for fulfilling *all* functions in the cluster *C*. That, I think, would still be too strong. For example, imagine an *extremely intense* instance of resentment, so intense that it is poorly suited for the purpose of initiating and facilitating a communicative process—it is so heated that it will drive its target away from being willing to exchange reasons with the blamer. But extremely intense resentment could still count as blame. As another example, imagine an *extremely servile* expression of resentment, so servile that it is poorly suited for marking out anything resembling a moral protest. But extremely servile criticism may still count as blame. My suggestion is that a blame-token only needs to be reasonably well suited for fulfilling *enough of the functions* in the function-cluster *C*. So an extremely intense instance of resentment could still count as blame if it has a robust

---

[16]I borrow these expressions—"well suited for," "good enough at," and "has what it takes"—from Vihvelin's (2013, 186–7) discussion of dispositions.

disposition to fulfill the protest and signaling functions, which it often does; and an extremely servile instance of criticism may still count as blame if it has a robust disposition to fulfill the communicative and signaling functions.

The suggestion here again can be compared to Boyd's notion of property-cluster. Important for Boyd, many particular things that belong to a homeostatic property cluster kind only exemplify some, but not all, properties in the relevant property-cluster. He acknowledges that it is sometimes difficult to settle whether a thing that only exemplifies a subset of the relevant property-cluster counts as a member. There will be some vague and indeterminate cases. A determinate answer, when there is one, would depend on various factors including "the causal efficacy of the partial cluster of properties" (Boyd 1988, 323). A token of a property-cluster kind is thus defined in terms of its exemplifying enough of the properties in the cluster. In a similar fashion, my proposal is that a token of blame is defined in terms of its robust disposition to serve enough of the functions in the cluster *C*. What counts as "enough" can be vague and indeterminate, but so is our concept of blame—otherwise, there would not be so many differing accounts regarding the exact extension of the phenomenon. However, vagueness and indeterminacy do not mean arbitrariness. To decide whether a token attitude/activity has what it takes to fulfill *enough* functions, it is important to examine both the *size* and the *value* of the partial function-cluster that it can potentially fulfill *in the relevant context*. Its size is simply about how much this partial function-cluster resembles the full function-cluster. This means that part of the definition is a number's game: the more functions in the cluster a token attitude/activity tends to fulfill, other things being equal, the more likely it counts as blame. But value is also important. If the fulfillment of this partial cluster of functions, in the relevant context, turns out to be valuable in a way comparable to the potential value of the full function-cluster, then this makes it more likely that the token attitude/activity counts as blame. Note that not any kind of value matters. If a partial cluster happens to have some aesthetical value in some context, then it does not affect whether the token belongs to blame-type. Instead, the value dimension asks us to examine whether and to what extent the partial function-cluster, in the

relevant context, makes a significant contribution to the realization of the aims of the system of first-order normative practices in an efficient and non-alienating way. That is, we would need to examine whether and to what extent this partial function-cluster's value resembles, in the relevant context, the potential value of the full function-cluster.

Let me illustrate the idea of blame-token functionalism using one example: dispassionate blame. Many have suggested that blame can be dispassionate in the sense that it lacks a feeling of emotional affect (e.g., Sher 2006), and even defenders of emotion-based accounts of blame take dispassionate blame to be a serious counterexample that they need to address (e.g., Tognazzini 2013; Menges 2017b). Sher describes the phenomenon as follows:

> We may, for example, feel no hostility toward the loved one whom we blame for failing to tell a sensitive acquaintance a hard truth, the criminal whom we blame for a burglary we read about in the newspaper, or the historical figure whom we blame for the misdeeds he performed long ago... blaming is something that we can do regretfully or dispassionately and that need not be accompanied by any rancor or withdrawal of good will. (2006, 88)

I am inclined to share Sher's intuition that these are genuinely instances of blame, but I am not sure. I suspect that many people have pre-theoretical intuitions like this; many of us are on a fence about whether these cases count as blame. Blame-token functionalism is well equipped to make sense of this indeterminacy in folk psychology, by explaining it in terms of indeterminacy about whether the relevant token attitude/activity has a robust disposition to serve enough of blame's functions. I am inclined to think that, for example, even when I feel no hostility or resentment toward the loved one, my condemnation toward them still serves important communicative and signaling purposes. Now, it may be the case that the lack of emotional heat makes my condemnation less of a protest, but it still seems to have a robust disposition to serve enough of the functions in blame's function-cluster. The partial function-cluster that my condemnation is robustly disposed to satisfy has both a decent size and a decent value. It has a decent size because it only lacks the protest function. It has a decent value because, if I were to communicate with my loved one and to signal my commitment to the relevant norm, then this

would make a significant contribution to the aims of the system of first-order normative practices in an efficient and non-alienating way. This process might be even more efficient if my blame could function as a protest, but its potential efficacy seems already high enough to resemble what we usually could expect from the realization of blame's full function-cluster. This is perhaps because, while in other contexts we may need protest, accompanied by the emotional heat of blame, to make for an efficient contribution to first-order aims, protest is not necessary in an intimate relationship like this. After all, romantic partners are often already quite ready to respond to each other's concerns and demands, so it is relatively easy to direct their attention toward us. I therefore think that this instance of dispassionate blame is genuinely blame. But many may disagree about all these different stages in the reasoning—they may think that dispassionate blame does not have a robust enough disposition to serve communicative purposes, or they may think that this partial function-cluster is not that valuable. They would therefore maintain that dispassionate blame, at least in this case, does not count as blame. The exact verdict is open to debate, but the cluster of function strategy offers a framework serviceable for those on different sides of the disagreement.

To summarize, my favored version of a pluralist function-based account involves both a definition of blame-type (according to its function-cluster) and a definition of blame-token (according to its robust disposition to serve enough of the functions in this cluster). Both definitions are crucial for the account to be appealing. I have presented protest, communication, and signaling as three different functions of blame. They are mutually reinforcing but cannot simply be derived from one another. These more substantial choice-points are important, but not as crucial as the more general framework that I have presented. Someone who thinks blame has more, fewer, or a different set of functions can fill in the cluster of function strategy in their own way.[17] This general framework is also compatible with thinking that there is *some* derivation among blame's different functions. For example, one may think that blame has an additional,

---

[17]For an example of an alternative set of functions see Cogley (2013), who argues that blaming emotions have an appraisal, a communicative, and a sanction function.

broad function of agency cultivation that is further derived from its protest, communicative, and signaling functions (see, e.g., Vargas 2013). The framework only requires that not all functions' roles can be derived from one single, unifying function of blame. It is thus even consistent with thinking that there is a paradigm function or a core function of blame—as long as it is paradigm or core in the sense that many, but not all, functions of blame are reduced to or derived from it.

3.4   Objections and Replies

3.4.1   Normative Depth?

One possible objection is a general worry about function-based approaches to blame: they are not informative when it comes to questions about the normativity of blame. More specifically, it does not follow from the fact that blame would fulfill its function to the normative conclusion that we *ought to* blame, or that blame is *appropriate* (see, e.g., Queloz 2021, 1368). This is sometimes described as a "wrong kind of reason" problem. For example, Hieronymi (2019) criticizes Fricker's (2016) function-based account by arguing that blame's functions are often disconnected from individuals' *reasons* to blame. We do not blame *because* doing so would fulfill certain functions. Using Hieronymi's analogy, a parent does not feel pride about her children *because* doing so would build her children's self-esteem, even if pride's social function concerns improving self-esteem (2019, 83).

My response is that a pluralist function-based definition of blame in fact has important implications about the blame's normativity. Questions about the potential gap between blame's functions and blame's normativity already came up in Chapter 2, where I suggested that a context-specific approach is promising in bridging the gap. My focus there is on how functions of blame can give rise to normative criteria about the *eligibility* problem: who are morally responsible agents or the appropriate targets of blame? Now that we have a fleshed out function-based view about the nature of blame, we can show that blame's functions also can give rise to normative

criteria about a distinct *value* problem: when and why is blame valuable? On my view, the definitive functions of blame are jointly valuable because they contribute to the aims of the system of first-order normative practices in an efficient and non-alienating way. Given that the relevant first-order norms (e.g., promoting morally right actions and cultivating better agential dispositions) are normatively valuable, my definition offers a promising explanation for why blame *as a type of practice* is normatively valuable. I also define an instance of blame in terms of its robust disposition to serve a certain number of functions. Again, this offers a promising explanation for why a *blame-token* is normatively valuable when it is, depending on the context-specific value of the partial function-cluster that the blame-token is disposed to fulfill. These are all still important normative implications of the view, though they by no means exhaust all questions about the normativity of blame.[18]

3.4.2   Theoretical Unity?

Another possible objection is that being a pluralist about blame's functions sacrifices theoretical unity. It may seem that we face an apparent dilemma: Either there is nothing that can unify blame's multiple functions, or there is something that can. If former, then it may appear that the list of blame's functions is arbitrary and ill-motivated. If latter, then it may appear that whatever unifies blame's multiple functions can simply be referred to as the single function of blame, and that should be what really figures in the definition of blame. Is there a way for a pluralist functionalist to satisfactorily solve this dilemma?

Now, my cluster of function strategy does involve something unified that we can say about blame's function-cluster: its constituents collectively contribute to the realization of various first-order normative aims in an efficient and non-alienating way. The list is thus not arbitrary or ill-motivated. However, this then leads to the other horn of the dilemma: why not just say

---

[18]I will come back to questions about the normative depth of functions, as well as the "wrong kind of reason" problem, later in Chapter 4. I will argue that function-claims have further implications about whether one ought to blame, in addition to their implications about the eligibility problem and the value problem.

that blame is defined by one function—the single, unifying function of contributing to various first-order normative aims in an efficient and non-alienating way? That, I think, would be overly generic. Nothing in my proposal entails that blame is the *only* kind that can make this sort of contribution. The definitive feature of blame lies in its making this sort of contribution *by employing its unique function-cluster*. The fact that blame has this function-cluster is partly explained by the value of this cluster in its contributing to the relevant first-order aims, but there is some degree of historical contingency that blame and its function-cluster happen to be selected or to persist as a paradigmatic or typical form of realizing these aims in our current practices. It may turn out that there are indeed better practices that we can endorse in order to fulfill the same aims in an even more efficient and even less alienating way. If such practices were to exist, they would not still be instances of blame but rather be what we should replace blaming practices with. This suggests that we cannot define blame using one single generic function. As an analogy, it is overly generic to claim that the function of the heart is to help keep a healthy body. A description about the heart's function is incomplete without reference to the particular mechanism that constitutes how the heart exactly contributes to bodily health. Likewise, my contention is that a description about blame's functions is incomplete without reference to the cluster of protest, communicative, and signaling functions.

A similar point can be made in response to someone who thinks that protest, communication (in the form of mutually exchanging reasons), and signaling all count as "communication" in some broad sense and maintains that blame has one single, unifying communicative function. One possible suggestion, for example, is that all three functions are unified by *communication for the sake of norm enforcement*. However, this will be both over-inclusive and under-inclusive. It is over-inclusive because not all forms of communication relevant to norm enforcement are blame's functions. For example, criminal punishment has a communicative role—it communicates to the wrongdoer that they have been proved to deserve the sentence (see Duff 2001). This form of communication is also for the sake of norm enforcement, but it does not belong to the function of

blame (despite being structurally similar; see Brink 2021). The suggestion is also under-inclusive. Blame's protest function, for example, is not merely for the sake of norm enforcement. It also has an important expressive and symbolic purpose, contributing to maintaining one's self-respect and to reaffirming one's moral beliefs. As I will discuss in more details later in Chapter 4, blame's protest function is primarily backward-looking, in contrast to its communicative or conversational function, which is primarily forward-looking. Therefore, we cannot unify blame's three functions under the label of *communication for the sake of norm enforcement*, which is primarily a forward-looking goal.[19] On my view, the description about blame's functions will be incomplete without reference to its particular cluster of protest, communicative (in the form of mutually exchanging reasons), and signaling functions.

3.5   Concluding Remarks

Existing function-based theorists often assume or write as if blame has a single, unifying function. I have provided an alternate approach: a functionalist can *both* accept the claim that blame has multiple functions *and* use these functions to define blame. It may initially seem that this pluralist alternative would have to rely on a collection of arbitrary, ill-motivated functions. But I have suggested that the cluster of function strategy is available for the rescue, and we can define blame in terms of a cluster of mutually reinforcing, jointly valuable functions. It may also seem that a function-based account is ill-suited for answering the question about whether a token attitude or token activity counts as blame. But I have suggested that we could offer an answer by examining the dispositional properties of the token. My pluralist function-based approach also fits well with the plausible claim that blame has multiple functions, as predicted by historical and ahistorical theories of functions alike. The upshot is that a pluralist function-based approach is

---

[19]In addition, blame may also have various cognitive and emotional functions, such as the function of *registering a true belief or thought*, that is also primarily backward-looking and distinct from the communicative function. I am inclined to think that this truth-seeking function is embedded in blame's protest function as well.

viable as a theory of blame, and it should be a serious candidate in the ongoing debate.

Chapter 3, in part, is currently being prepared for submission for publication of the material (as "Blame: A Pluralist Function-Based Approach," by Wang, Shawn Tinghao). The dissertation/thesis author was the primary investigator and author of this paper.

Chapter 4

Blame Conflict

4.1 Introduction

In some circumstances, one is faced with conflicting considerations with respect to whether to blame an individual. Call this a "Blame Conflict." Consider the following example:

> *Unjust System.* Yue is a victim of racial injustice. She is aware of the wrongdoings done by the perpetuators of racial injustice in her society and is inclined to blame them. However, Yue lives in a non-ideal society such that the attempt to communicate with these perpetuators is highly likely to fail, and indeed likely to backfire. As Yue knows, if she blames the perpetuators, it is likely that her blame will turn out to be dismissed by the perpetuators, to alienate potential allies, and to help very little with the fight against racial injustice. But if she does not blame the perpetuators, it is likely that she will draw more allies and contribute more to the fight against racial injustice.[1]

It is unsurprising that Yue will feel torn. She is subject to serious injustice, and she thereby seems justified to issue a complaint or a protest. However, the fact that blame would not be listened to by the perpetuators, and would alienate potential allies, is a consideration against blame. Cases like this, I think, constitute an important kind of practical conflict. My goal in this chapter is (i) to provide a novel account of why this kind of conflict occurs, and (ii) to propose

---

[1]For similar examples about whether one ought to feel anger about social injustice, see Srinivasan (2018).

a theoretical framework that illuminates the various kinds of factors that should be taken into consideration in the resolution of many of these conflicts. More specifically, I will apply the pluralist function-based approach to blame in my analysis of the phenomenon, and characterize the conflicting considerations in Blame Conflict, or at least a theoretically interesting kind of Blame Conflict, in terms of blame's conflicting functions.

4.2   A More Accurate Characterization of Blame Conflict

On my definition, a person faces a Blame Conflict if and only if

(1) There is a consideration *R1* that speaks in favor of an appraiser blaming some target individual for an action,

(2) There is a consideration *R2* that speaks against the appraiser blaming the target for the same action, and

(3) Neither *R1* nor *R2* ceases to be a consideration for her just because it conflicts with the other consideration.

What does it mean to say that there is a consideration for her that speaks in favor of or against blaming someone? Several points are worth making.

First, my primary focus is on *overt blame* or blame as a *blaming activity*, typically in the form of an outward behavior or speech act that expresses one's negative reactive attitudes like resentment, indignation, and guilt. There is a further issue regarding the potential conflicting considerations about *covert blame* or whether to hold a *blaming attitude* like resentment, indignation, and guilt to begin with. Some believe that conflicts about blaming activity and conflicts about blaming attitudes are closely tied to each other. Srinivasan, for example, thinks that anger is partly constituted by the stereotypical expressions of anger, in the absence of deliberate cultural training (2018, 137–8). One may hold the same view about those angry emotions involved in a blaming attitude, such as resentment, indignation, and guilt.[2] But some would disagree. Hieronymi (2019), for example, points out that blaming attitudes are non-voluntary such that they cannot be assessed

---

[2]See, e.g., Fricker (2016, 2018).

in the same way as voluntary blaming activities. According to Hieronymi, the ethics of blaming attitudes is sensitive to a much narrower set of considerations compared to the ethics of blaming activities. I will be neutral between these two opposing positions and focus on Blame Conflict in the case of overt blame or blaming activities.

Second, there is a consideration *R* for an individual *S* that *speaks in favor of* blaming just in case that the following conditional is true: If *R* were to be the only relevant consideration for *S* that bears on the question of whether to blame, then it would be the case that *S* ought to blame, all-things-considered. Similarly, *R* speaks against blaming for *S* just in case that, if *R* were to be the only relevant consideration, then it would be the case that *S* ought not to blame, all-things considered. Considerations, on this picture, are facts or purported facts that bear on a question.[3] This relation between individual considerations and the all-things-considered-ought is structurally like Ross's distinction between pro tanto duties and all-things-considered duty. Different considerations pull an individual into different directions, and it is impossible for the individual to comply with all the considerations when she faces a practical conflict.

Third, my focus is on whether someone all-things-considered ought to blame an individual for an action (in the form of a blaming activity), and this is broader than the question of whether the individual is *blameworthy* for the action.[4] The former question is sometimes referred to as a question about the "ethics of blame" (Scanlon 2008, 123; Coates & Tognazzini 2013, 17). Coates and Tognazzini, for example, suggest that the ethics of blame covers not only conditions

---

[3]The relationship between a consideration and a reason is controversial. Some think that considerations are reasons themselves (e.g., Raz, 1999), but others think that reasons are instead mental states that have considerations as their propositional contents (e.g., Smith, 1994). I do not wish to take a stand on this issue and will stick with the formulation of Blame Conflict using considerations rather than reasons. Note that there is a sense of "speaking in favor" in which *R* can speak in favor of an action without giving rise to an "ought"-claim, even if there are no other relevant considerations. Considerations like this concern "enticing reasons" rather than "requiring reasons"; however, the idea of merely enticing reasons is controversial (see, e.g., Raz 1999; Dancy 2004). I will only focus on those considerations that concern requiring reasons and could give rise to "ought"-claims. A related distinction is between an "ought"-claim as a "specific action demand" and as an "axiological recommendation" (Pereboom 2014, 141). I will focus on considerations that give rise to specific action demands rather than merely a recommendation to the effect that it would be good for someone to blame.

[4]Some think that blameworthiness implies a pro tanto reason to blame, but not the other way around (e.g., Brink 2021, §§39, 41, Brink & Nelkin 2022).

of blameworthiness (conditions that the transgressor need to satisfy), but also what they call "the conditions of jurisdiction"—those conditions that the would-be blamer needs to satisfy, and "the conditions of procedure"—those conditions that the blaming interaction needs to satisfy (2013, 17–8). However, it is worth noting again that we should pay special care to distinguish between norms about the ethics of blaming attitudes and the ethics of blaming activities—it is possible that they have different conditions of jurisdiction and conditions of procedure. My primary focus is again on the ethics of blaming activities.

4.3   Blame and Its Functions

My proposal is that many cases of Blame Conflict can be understood and illuminated by attending to the *multiple functions* that blame has. This proposal is meant to be an application of the pluralist function-based approach to blame defended in Chapter 3. In particular, I will focus on conflicts between blame's protest and communicative functions. However, I do not think that one needs to accept the specific account I defended there in order to accept the function-based characterization of Blame Conflict that I shall advocate in this chapter. Instead, one only needs to be convinced that blame has multiple functions—more specifically, a backward-looking function of moral protest and a forward-looking function of moral communication. In this section, I will focus on explaining and motivating this claim.

There are many differing accounts of what it is for something to be the function of an entity-type. I will briefly describe two of the most influential accounts. On one view, *F* is a function of *X* if and only if (i) enough of past tokens of *X* did *F*, and (ii) the fact that (i) is true in part explains why current tokens of *X* exist now (e.g., Wright 1973, 1976). This is the *etiology theory of functions*.[5] On a different view, *F* is a function of *X* if and only if (i) *X* has a

---

[5]In biology, this etiological history is often taken to be a history of natural selection, and *F* is meant to indicate a selective advantage of *X*. However, I think that the general point of the etiology theory can be made without assuming that the etiological history only involves natural selection. Instead, it can involve the broader history about the causal or explanatory basis for a certain kind of practices that we currently have.

disposition to do *F*, and (ii) the disposition described in (i) contributes to the proper functioning of a broader system that includes *X* as a subsystem (e.g., Cummins 1975). This is the *system theory of functions*. The two views will, for example, agree that it is a function of the human heart to pump blood, but disagree about why. According to the etiology theory, it is because many past tokens of the heart have done the job of blood-pumping, and this evolutionarily explains why we humans currently have hearts now. By contrast, the system theory justifies blood-pumping as being the function of the heart by appealing to the current disposition of human hearts to pump blood and how this disposition can contribute to various capacities of a human circulating system, such as the capacity to deliver oxygen, nutrition, and so on.

What, then, is blame's function? I take that blame has at least two functions: one backward-looking and the other forward-looking. The backward-looking function involves moral protest. Smith, a defender of moral protest theory of blame, characterizes blame's protest as consisting of two parts: the registration of a truth that someone is culpable of some wrongful conduct, and "a challenge to, and a repudiation of" the moral claim implicit in this wrongful conduct (2013, 42). Moral protests are primarily backward-looking since they are, after all, protests of *past culpable wrongdoings*. Whether a protest is ultimately successful or brings out desirable consequences can depend on various sorts of future-directed facts, but whether an instance of blame is *fitting* for its moral protest function is largely a backward-looking matter, depending on whether the protest registers a truth and marks out the wrongful aspects of a conduct that is worth challenging.

One might assume that a wrongful conduct always implies a false moral *claim*. However, if not all wrongs are wrongs of disrespect, then not all moral protests are protests of false moral claims. It seems that we can protest against a substandard *moral treatment* when, for example, someone violates an obligation that she owes to us, even when this obligation-violation does not imply any false moral claims. I will therefore put the protest function of blame more broadly: it consists in *registering and challenging a culpable wrongful conduct*, where one either challenges

a false *moral claim* or a substandard *moral treatment* that is implicated in the conduct.

But blame also has another, more forward-looking, communicative or conversational function. While Smith takes the moral protest function to be definitive of blame, she also points out that blame has another function: it "implicitly seeks some kind of moral acknowledgment on the part of the blameworthy agent and/or on the part of others in the moral community" (2013, 43). The thought is that blame does not merely express a unilateral protest, but also seeks uptake through a bilateral form of *communication*. In blaming someone, we do not merely aim to register and challenge about what has been done, but also aim at eliciting the right kind of reactions and responses. This second aim is primarily forward-looking. Whether blame is *fitting* for its communicative function, I think, largely depends on whether its target is someone that has the capacity to engage in this communicative process, a capacity to be exercised in the *future*. While Smith's characterization of the forward-looking function of blame focuses more on eliciting *acknowledgement* of the wrongdoer, McGeer (2012, 2013, 2019) maintains that an educative or transformational aspect is also an essential part of this function. Blame does not merely seek acknowledgement; it seeks acknowledgement *and* seeks to thereby effect a change in the wrongdoer's future attitudes or behaviors. Funk, McGeer, and Gollwitzer (2014) highlight this point and argue that empirical evidence speaks in favor of blame as aiming at not just "sending a message"—it also aims at "effecting a change in the offender's blameworthy attitudes and/or behavior" (987–8). Though Smith and McGeer disagree about what kinds of uptake blame seeks through communication, they share the view that moral communication is central for blame's forward-looking function, and that this function involves prompting or priming the wrongdoer's future attitudes or actions in a certain way. In what follows, I will put it that the forward-looking function of blame involves *prompting better attitudes/behaviors through moral communication*.

Note that while Smith seems to imply that the backward-looking and forward-looking aims are two distinct functions of blame, McGeer tends to put it that they are two *dimensions* of blame's function. One may accept this latter claim and insist that blame only has one function,

and the two aims described above are merely different elements or dimensions of this same function. I find the former approach more appealing for a couple of reasons. First, thinking of blame as having more than one function better captures our intuition about cases where the backward-looking and forward-looking aims come apart. For example, consider cases where blame is counterproductive in prompting better attitudes/behaviors through moral communication, but it still protests a culpable wrongdoing. It seems more intuitive to claim that blame then fails to achieve *one of its functions*, instead of claiming that blame does not achieve any function of the practice at all. Second, both the goal of protesting a culpable wrongdoing and the goal of prompting better attitudes/behaviors through moral communication seem to satisfy the conditions often deemed necessary for something to count as blame's function. Consider the etiology theory of functions first. Past tokens of blame did result in both protesting culpable wrongdoings and prompting better attitudes/behaviors. Both historical facts also seem to play a role in the explanation for why we still have the practice of blame now. Macnamara has argued that blame's function of prompting acknowledgment of fault, particularly in the form of eliciting guilt, has historically contributed to the "building and healing of community" (2015b, 560). Without acknowledging our faults, it seems unlikely that we can come to "caring about moral values" and to have a "susceptibility to the reactive attitudes" now (Macnamara 2015b, 561). I take the same sort of argument, however, as also supporting moral protest as being a function of blame. After all, in the absence of our protesting against culpable wrongdoings at past, it also seems unlikely that we can come to care about moral values and be emotionally susceptible to the reactive attitudes now. It thus appears that both the backward-looking and the forward-looking aims should count as functions of blame. The same conclusion also seems to follow from the system theory of functions. One way to understand the relevant larger "system" is in terms of a *broad system of our moral practices in general*, which may further include moral behaviors, attitudes, norms, norm enforcements, and their interlocking relations with each other. The two aims of blame provide distinct contributions to various capacities of this broad system. For example, effective

norm enforcement seems inadequate without our constantly protesting against others' culpable wrongdoings, and agency cultivation and moral education seem inadequate without our being able to recognize our past wrongs through communicating with those we have wronged. Again, it is hard to see why we cannot treat the two aims as two distinct functions of blame.

The thought that blame has more than one function is not anomalous. A lot of entities, mechanisms, and practices have more than one function, and sometimes those functions can come into conflict with each other. Importantly, many practices in the vicinity of blame appear to have multiple functions. For instance, punishment seems to both have a backward-looking, retributive function and a forward-looking, deterrence function. Forgiveness has a backward-looking function of acknowledging amendatory attitudes as well as a forward-looking function of restoring one's relationships with the person who apologizes. My contention is that a similar pluralist framework can be used to understand blame: it has both a backward-looking and a forward-looking function, which may conflict with each other in certain contexts.

4.4   Why Blame Conflict Occurs

Now that I have sketched a pluralist function-based framework about blame and explained its appeal, we can supply a novel explanation of why a theoretically interesting kind of Blame Conflict occurs. My contention is that this kind of practical conflict occurs because of the conflicts between blame's different functions. I will first describe how the pluralist function-based account explains this kind of conflict, and then explain what is special about these conflicts compared to other situations where we face conflicting considerations about whether to overtly blame.

The forward-looking function of blame, I have suggested, is to prompt better attitudes or behaviors through communication. In *Unjust System*, blame is unlikely to realize this function due to Yue's non-ideal situation. Though Yue can initiate a communicative process, it is unlikely that her social community will be willing to engage with her. The wrongdoers are unlikely to

acknowledge that their racist attitudes and behaviors violate moral norms, and the third parties are unlikely to engage with Yue in a way that helps with improving her situation. In fact, Yue's blame is likely to backfire, since, as we stipulated in the case, blame has an alienating effect in her society and can overall decrease the likelihood for the target of blame to form better attitudes or behaviors in the future.

However, Yue's blame can still realize the backward-looking function of the practice: to protest against a wrongful conduct. Though blame, Yue expresses her challenge and repudiation of a false moral claim or a substandard moral treatment as implicated in a conduct. It is the case that blame's protest function often is linked with its communicative function, and in the best scenario it achieves both functions at the same time. But the mere fact that blame cannot satisfy its forward-looking function does not entirely undermine the positive values of moral protest. Among other things, protesting against wrongful conducts can promote one's self-respect, reaffirm one's moral beliefs, signal one's commitment to a norm, and draw people's attention to an issue that needs to be addressed (see, e.g., Boxill 1976; Talbert 2012). Now, Yue's society may be so non-ideal such that it silences the voice of its minorities in a way that blame cannot even amount to a publicly recognized protest. It would seem that blame cannot even fulfill its protest function. But even then, we may still think that Yue's blame could count as a form of *private protest*: Yue protests in her mind, and she takes her blame to be a form of protest in her own sight. A private protest may not be as valuable as a publicly recognized protest; nonetheless, it still has positive values, such as promoting self-respect and reaffirming one's moral beliefs.

A theoretically interesting kind of Blame Conflict occurs, then, when the protest function and the communicative function of blame cannot be achieved at the same time. But there is more to say about this conflict. To start with, I think that blame's functions, when jointly realized, have a distinctive value of *making us better moral agents in an efficient and non-alienating way*.[6]

---

[6]For related discussion on blame and agency cultivation, see, e.g., Vargas (2013) and McGeer (2019). The thought here is a more concrete version of my earlier claim in Chapter 3, where I suggest that the joint value of blame's functions involves contributing to first-order normative aims in an efficient and non-alienating way. Here I focus on making better agents as a valuable first-order moral aim in the context of Blame Conflict.

What does this mean? On one hand, blame is quite efficient: it at the same time signifies a unilateral protest against a culpable wrongdoing and initiates an interactional communicative process, and it often makes the wrongdoer feel guilty and painful in a way that motivates them to be a better agent. This is why the transformative force of blame tends to be stronger than some nearby practices, such as grief and hurt feelings, which are much less likely than blame to produce a practical impact on the wrongdoer. Empirical evidence strongly supports that blame is an efficient practice when it comes to these goals (for discussion and a comprehensive review see Clark forthcoming). When it comes to self-blame, there is also plenty of evidence suggesting that guilt is efficient: it is shown to be associated with "a corrective action that an individual can take (but does not necessarily take) to repair the failure" (Lewis 2008, 748). I take these to be sufficient to establish the general claim regarding the efficacy of blame, though *how efficient* blame is in realizing the relevant aims deserves further investigation.[7]

On the other hand, blame is also *non-alienating* in a way similar to emotional attitudes like hurt feelings. This is what distinguishes it from punishment, scolding, or threats, which all have a notoriously alienating effect. I take it that blame, as a type of practice, is non-alienating in at least three respects. First, blame is less likely to *alienate the blamer from the wrongdoer*, in the sense of causing a separation between the two people to an extent that makes certain relationships impossible. Even close friends or close family members do blame each other in ways that do not seem to impair their relationship (see Wolf 2011). By contrast, if we treat a close friend or a family member with threats or punishment, then these practices are more likely than blame to undermine the relationship. I do not deny that sometimes even blame can significantly damage a relationship, but the claim here is meant to be a generic claim that compares different kinds of practices. Second, we can understand the non-alienating power of blame using Strawson's (1962) distinction between the *participant stance* and the *objective stance*. When we blame individuals, we still treat them as morally responsible agents and people that we can *engage with*. In this

---

[7]Another issue is if there are similarly efficient practices for the same aims, such as moral education. I will put this aside and revisit this issue in Chapter 7.

sense, we still treat them as *participants* of the moral community, and thus we are not alienating them away from the participant stance. By contrast, when we threaten or warn someone, we already get closer to an objective stance that treats our target as not a morally responsible agent we can engage with, but someone "to be managed or handled or cured or trained" (Strawson 1962, 8). Third, when we blame someone, the target of blame is less likely to feel that *she herself is alienated from the function of the practice*. When we threaten someone or punish someone, for example, our target does not herself *actively participate* in the process of realizing the functions of the relevant practices. They are merely forced to perform certain actions due to the force of the threat or punishment. By contrast, blame seeks to make us better agents by situating us in an emotionally invested sort of practice. This is best reflected in blame's communicative function. The wrongdoer is invited to together engage in the relevant communicative process, and she is thus invited to actively participate in the process of realizing blame's functions and values. Thus, the participants of blaming activities are more likely (than, for instance, the participants of punishing activities) to see themselves as having a *personal interest* in the goals of blame and to feel that they have *personally participated* in the process of reaching these goals. Blame's protest and communicative functions jointly amount to this efficient and non-alienating agency cultivation process.

Because the joint realization of blame's functions has this distinctive value, when the joint realization is impossible or highly unlikely, as in the kind of Blame Conflict I highlighted, there is not just the danger of failing to achieve one of blame's functions, but also the danger of losing the possibility of cultivating agents in an efficient and non-alienating way in the relevant kind of contexts. This is even more worrisome when the conflict is due to a *systematic disruptor* of blame's normal process of realizing its both aims. *Unjust System* can be a case like this. Why does Yue's community not listen to her blame? One possibility is widespread testimonial injustice regarding Yue's social subgroup. Another possibility is that there are "echo chambers" or "epistemic bubbles" in the social epistemic structures of her society (see, e.g., Nguyen, 2020).

These social epistemic structures can block people outside Yue's chamber or bubble to hear her complaint in a fair and charitable way. Both explanations concern a fairly general mechanism that tends to systematically disrupt the joint realization of blame's functions when blame is issued by a member in Yue's social subgroup, and toward a member in the more privileged subgroup in her society.

The upshot is that, as a general matter, Yue's group systematically lacks the possibility of using blame as a tool to motivate people in the privileged social group to be better agents in an efficient and non-alienating way. What would happen? It is natural to think that they then would *either* choose the efficient but alienating route by treating people in the opposing group with cold punishment, scolding, and threats, *or* choose the non-alienating but non-efficient route by expressing their dissatisfaction through sadness and disappointment. We risk either causing even more alienation and hostility in the society, or causing more injustice to the already oppressed group in Yue's society. This is one reason why I find cases like *Unjust System* constitute an especially interesting kind of Blame Conflict. It indicates both a local failure of blame's operating system and a risk of a more global, systematic failure: blame's failure to make better agents in an efficient and non-alienating way.

Another reason why cases like *Unjust System* are special is this: they are cases where the counterproductivity critique of blame is, at least intuitively, especially pressing and hard to respond. This is the critique that we often hear when we feel and express our resentment or indignation: we should not feel this way or express these feelings because these angry emotions are "counterproductive," "useless," or "do not fulfill any purpose."[8] However, if the critique is merely about counterproductivity *per se*, it is mysterious why this line of reasoning looks appealing at all. Counterproductivity alone does not constitute a valid *general* argument against a behavior. Blame in particular can be *apt* or *fitting* even when it does not bring out the most desirable consequences, a point convincingly argued for by Srinivasan (2018).

---

[8]For a version of the critique, see Nussbaum (2016).

However, I do not think that the counterproductivity critique can be addressed simply by distinguishing between productivity and aptness. This is because there is at least one sense in which whether blame is *apt* or *fitting* does seem to depend on its consequences—aptness or fittingness with regard to the forward-looking function of the practice. In contexts where the counterproductivity critique looks most pressing and hard to respond, it is not just about any kind of counterproductivity, but *counterproductivity regarding the very forward-looking function of blame*. The critique is that our blame is counterproductive in those very forward-looking aspects—on my view, prompting better behaviors/attitudes through communication—that our practice of blame is "designed to" or "built to" bring about positive consequences in. It is thus unsurprising that we would feel that this kind of counterproductivity is a pressing concern: it makes us wonder if blame defeats its very purpose, and if blame has any point at all. The problem with the counterproductivity critique is thus not just an obvious fallacy. But it does overlook something very important: the forward-looking function of blame is not the practice's *only function*. Blame also has the backward-looking function of moral protest, and blame can realize this function even when it cannot achieve the forward-looking goal. We thus have a diagnosis of the counterproductivity critique that both explains its initial appeal and supplies a reasonable response.

I have motivated and defended an explanation for why a theoretically interesting kind of Blame Conflict occurs: because blame cannot at the same time achieve its multiple functions. I have argued why this kind of Blame Conflict is especially interesting: it indicates both a local failure and a risk of a global failure for blame to make better agents in an efficient and non-alienating way, and it explains why the counterproductivity critique of blame is hard to respond. The scope of my thesis, however, should not be overstated. There are surely other also important kinds of Blame Conflict that are excluded from the pluralist function-based analysis. For example, sometimes there are considerations against blaming simply because the blamer is under personal danger of various sorts, such as when a woman is afraid of retaliation from the

harasser. Cases like this point to issues that need to be addressed, and they are sometimes in no way less urgent than the ones pointed to by the kind of Blame Conflict under the pluralist function-based analysis. My analysis thus is best described as an account of one (instead of the only) theoretically important kind of Blame Conflict.

4.5 The Normativity of Blame's Functions

It is worth noting that my account of Blame Conflict makes sense only if blame's functions bear normative significance. They must give rise to, or in some other way related to, considerations that speak in favor of or against blaming. But the inference from functions to normative considerations is controversial. Why should we think that whether blame fulfills a certain function is a normative consideration for us regarding whether we *ought to* blame? Answering this question would require us to explore the normative implications of blame's functions, which I will try to address in the current section.[9]

There are two ways to challenge the inference from blame's functions to normative considerations about whether one ought to blame: a challenge from the morally perverse functions, and a challenge related to the "wrong kind of reason" problem. I will address them in order.

4.5.1 Morally Perverse Functions?

It might be suggested that blame can have morally perverse functions. In addition to issuing protests and initiating moral communication, it might be suggested, blame can also have the function of protecting in-groups and excluding out-groups, or the function of reinforcing existing social hierarchy and marginalizing the social minorities. But it seems that neither the excluding-out-groups function nor the marginalization function should impact whether we *ought*

---

[9]Note that this question—whether we ought to blame—may come apart from (despite closely related to) other questions about the normativity of blame that I have addressed earlier, such as the question of who are blame's eligible targets (Chapter 2) and the question about the value of blame (Chapter 3).

to blame. Even if blame would fulfill its marginalization function, this does not mean that we ought to blame—quite the opposite, it seems that we ought not to blame because of the perverse nature of the marginalization function.

But we should be careful when attributing these perverse functions to blame as *a type of practice*. It is obvious that some *tokens* of blame bring out these negative consequences, and it is also quite plausible that some *subtypes* of blame—for example, the subtype of blame sometimes called "blaming the victim"—may have these functions. But the functions of blame's subtypes can come apart from the functions of blame-type. Note that the operations of both the excluding-out-groups function and the marginalization function seem to depend on some socially contingent factors. The former only operates when there is a sharp divide among social groups in a society, and the latter only operates in a non-ideal social hierarchical system. It is thus quite plausible that they are only functions of those subtypes of blame that occur in the relevant non-ideal social conditions. By contrast, the protest function and the communicative function that I have identified do not seem to be so socially contingent. I thus find it more plausible that these are functions of blame as a general type of the practice.

What if it turned out that blame does have some morally perverse functions? I am inclined to agree, then, these morally perverse functions do not have much *positive* normative significance— it does not follow from the fact that blame would fulfill them that we ought to blame. But it seems to me quite plausible that these functions would then have *negative* normative implications: they provide considerations that speak against blaming. If I learned that my blaming would contribute to the function of excluding out-groups, then I have a pro tanto reason against blaming. Therefore, the upshot from morally perverse functions is not that functions do not matter. Instead, the lesson we should draw is that blame's functions do have normative implications, but whether they have positive or negative implications depend on the nature of a function. Cummins's system theory of function will be helpful here. Remember that, according to this theory, blame's function is defined by what blame does that contributes to the proper functioning of a broader system. My

suggestion then it the following: it is only when a function of blame contributes to a *morally good* broader system that this function has a *positive* normative implication about whether we ought to blame. Suppose, for the sake of argument, that blame has a marginalization function. According to the system theory, this means that (i) blame has a disposition to marginalize people, and (ii) this disposition is appealed to in the adequate and appropriate explanation of the proper functioning of some broader system. It is quite plausible that this broader system is morally perverse—for example, a system of social oppression. Because this is a morally perverse system, the functions that contribute to the system have negative normative implications. By contrast, blame's protest and communication functions do carry with positive normative implications because they contribute to a morally good system: the system that involves moral practices in general, with the internal aim of cultivating better agents in an efficient and non-alienating way.

It might be objected that what is doing the real theoretical work is the *moral quality* of the functions, instead of functions themselves. But note that the apparent moral quality of a function can come apart from the moral quality of the broader system that it contributes to. Imagine that blame has an apparently innocent function but it contributes to a morally perverse system. This function would still have negative normative implications. Functions are thus doing real work in my account of Blame Conflict, partly by directing questions about the normativity of blame toward questions about the moral quality of the broader systems that these functions contribute to. The appeal to functions thus remains central in my account of Blame Conflict.

4.5.2   Wrong Kind of Reasons?

There is another challenge against the function-based explanation of Blame Conflict. Even if blame would fulfill some morally good function, it might be suggested, this still does not give rise to the claim that one *ought to* blame. In particular, the worry is that functions often provide only *wrong kinds of reasons* to blame, which can come apart from those reasons that actually motivate us to blame. This line of thought is part of Strawson's (1962) criticism of what he calls

the "optimist"—theorists who understand blame in terms of its social regulative function. As Strawson puts it, "to speak in terms of social utility alone is to leave out something vital in our conception of these practices" (1962, 24). The thought is that social utility and social function are both wrong kinds of reasons to blame. Hieronymi (2019) raises the wrong-kinds-of-reason challenge against Fricker's (2016) communicative function-based account of blame. One of Hieronymi's examples involves the nearby phenomenon of pride: even if the social function of pride concerns building self-esteem, "I do not—and could not—feel great confidence in them [my children] or take pride in their accomplishments in order to build their lagging self-esteem" (2019, 69). Again, the worry is that a function-based account of the phenomenon is leaving out something crucial in our phenomenology and in our motivating reasons. Just like a function-based account of blame leaves out something vital about blame, it may be suggested that my function-based account of Blame Conflict leaves out something vital about the nature of such conflicts as well.

In reply, note first that it is unclear if the wrong kind of reason problem applies to *blaming activities*, which are the primary object of concern in this chapter. Hieronymi, for example, emphasizes that the objection is primarily about non-voluntary blame—those *mental attitudes* that constitute blame. Importantly, it seems that we have little to no direct control about whether to have these attitudes. Even if we know that having an attitude has a positive function, it does not follow that we have a reason to have that attitude, because it may be a reason that we just cannot form an attitude with. In contrast, the scope of reasons that can be taken into consideration in deciding whether to perform an action is much broader. This is admittedly controversial, and some theorists maintain that a similar right and wrong kind of reason distinction applies to many behaviors and activities as well (see, e.g., Schroeder 2010).

Even if there is indeed a wrong kind of reason problem in the case of blaming activities, I take it that there is a modification of my account that can avoid the objection. The idea is to combine the fact that blame would fulfill a certain function with an individual's *adopting that*

function as one of her ends.[10] Yue, for example, can adopt both blame's protest function and communicative function as her ends, and there is nothing irrational or psychologically unrealistic for her to do so. By contrast, it seems that it is psychologically unrealistic for a parent to adopt pride's self-esteem building function as her end—it is unrealistic, at least for many, to think that we can be sincerely proud of our children while at the same time aiming at using pride to improve their self-esteem. What explains this difference? My tentative suggestion is that pride's self-esteem building function is quite alienating. It is thus often psychologically unrealistic for us to participate in a personally and emotionally invested sort of practice, like pride, while at the same time adopting an alienating end as the goal of our practice. However, blame's functions, when jointly realized, have a value of contributing to the system of moral practices in an efficient and non-alienating way. The functional system of blame is built in such a way that we can fairly easily adopt the relevant functions as our ends when we blame. At the least, there is typically nothing irrational or psychologically unrealistic for us to do so.

To summarize, I considered the objection that my account of Blame Conflict relies on the following, controversial assumption:

> *Function-Reason Link*. If one function of blame is to *F*, then there is a consideration for an individual *S* that speaks in favor of blaming when blame would fulfill this function.

I have argued that, while the Function-Reason Link faces serious challenges, a modified version of the claim can counter these challenges:

> *Function-Reason Link**. If (i) one function of blame is to *F*, and (ii) *F* contributes to a morally good broader system, and (iii) *S* adopts *F* as one of her ends, then there is a consideration for *S* that speaks in favor of blaming when blame would fulfill this function.

---

[10]Something like this has been suggested by some authors. For example, Hieronymi writes that we may be "taken up, in some way, into" a social function (2019, 77). And McKenna, uses the analogy to linguistic conversations, puts it that we can design our actions to "exploit the conventions others holding responsible use for interpreting behavior" (2012, 87).

84

4.6   How to Resolve Blame Conflicts

I have identified a theoretically interesting kind of Blame Conflict and offered a pluralist function-based account of why it occurs. The next question to address is whether one *ought to blame* if such an account is on the right track. If blame would realize its backward-looking function but at the same time undermine its forward-looking function, which of the two functions should we prioritize? This section aims to employ the pluralist function-based analysis to discuss two kinds of factors that we should consider in order to answer this question: the general significance and the context-specific significance between blame's functions.

4.6.1   General Significance

The first question one may ask is if one of blame's functions is *generally more significant* than the other. If so, then this suggests that we should weigh the failure to achieve these functions differently in our deliberative processes. Note that blame's protest function is more concerned with the *blamer*, whereas its communicative function is more concerned with the *blamee*. Theorists have taken different sides on the matter of whether the ethics of blame is more about the blamer or about the blamee. Some theorists (e.g., Russell 2004; McKenna 2012; Fricker 2016) argue that the propriety condition of blame primarily depends on the various features of the blamee, such as whether the blamee has the relevant emotional, conversational, and communicative capacities. But others (e.g., Talbert 2012; Hieronymi 2019) have argued that, in answering the question of whether blame is appropriate, we should place a priority on the needs and interests of the blamer. This debate has obvious implications on the question at issue—those who take the perspective of the blamer to be more important is more likely to prioritize its protest function, whereas those who treat the perspective of the blamee as more primary is more likely to prioritize its communicative function.

Instead of choosing a side in this debate, I am inclined to endorse a symmetrical form

85

of pluralism that treats the perspectives of the blamee and the blamer as equally or similarly significant for the purpose of deliberating about whether to blame. Blame Conflict is pressing exactly because it exemplifies the tension between these two perspectives. If one perspective were to be generally more primary than the other, then we would feel not ambivalence, but only regret, in this kind of conflicts. I am not able to give a full defense of this no-priority position here. Instead, my approach is to motivate its plausibility by discussing and challenging some alleged arguments for the general primacy of one of blame's functions.

First consider the position that the protest function of blame is more significant. One may argue that blame's forward-looking function is only successful when blame also realizes its backward-looking function of protesting. It may be suggested that, for example, blame does not just aim at communicating, but aims at communicating about a false moral claim or a substandard moral treatment implicated in one's conduct. And blame's communicative function is achieved only if it also registers and challenges a claim or a treatment that is fitting to register and worth challenging. The implication is that blame's protest function is more *fundamental* than its communicative function. Shouldn't this imply that blame's protest function is one that we should place more priority on?

In response, I share the intuition that, in order for blame to successfully prompt better behaviors/attitudes through communication, the communicative process must be always based on our successfully registering and challenging what is worth registering and challenging in the first place. That is, it may seem that realizing blame's forward-looking function *requires* its backward-looking function to be realized, but not the other way around. However, the apparent force of this asymmetry is, I suspect, based on an overly ideal picture of how moral communication works. Real-life scenarios contain cases where successful communication may require compromising truth. Calhoun (1989, 1999, 2016) has convincingly argued for a point like this. In discussing the aim of communication in our moral life, Calhoun writes that,

> . . . the highly discursive nature of philosophical practice works to obscure how problematic real-life moral communication may be. Making one's choices legible,

and one's reasons intelligible as justifying reasons, is much more likely of success when choices and reasons can be carefully formulated in essays and books. As a philosopher, for example, Claudia Card can devote an entire chapter to articulating what she means by "lesbian" and thus what she is morally up to when she stands up for her lesbian life. As a participant in the daily practice of morality, she does not have this luxury. As a result, exercises of integrity, like hers, that involve rejecting conventional meanings are bound to be and will remain illegible to others as an exercise of integrity or even as minimally justified. (2016, 45)

Calhoun's point is that real-life moral communication faces more struggles and roadblocks than communication among moral philosophers. When moral philosophers engage with each other, they usually have the luxury of articulating the truth that they take to be true and providing detailed justification for it. But agents undertaking ordinary moral communication often do not have this privilege. The upshot is that they are more likely to see a clash between two ideals: what Calhoun calls the moral ideal of "getting it right" (2016, 42) on one hand, which consists of searching for correct principles and adequate justifications, and the moral ideal of "participating in a shared scheme of social cooperation" (2016, 42) on the other, which consists of our communicating our moral views to others, seeking consensus, etc.

My suggestion is that a similar clash of ideals is manifested in Blame Conflict, but here it is in the form of conflicting functional goals. Thinking abstractly, it may appear that blame's communicative function is preconditioned by its protest function. However, sometimes successful moral communication requires sacrificing, at least to some degree, the extent to which blame *accurately* registers or challenges a false moral claim. That is, we may sometimes find it easier to communicate if we blame a person *to a lesser degree* for what she has done than she deserves, or blame her for *lesser things* than she is blameworthy for. For instance, Yue's blame seems more likely to prompt better behaviors/attitudes through communication if she somewhat downplays how wrong her moral interlocutors are in perpetuating racial injustice. This kind of cases suggests that, at the least, the best realization of blame's forward-looking function does not always require the best realization of blame's backward-looking function. It may still be the case that we need some truth or accuracy condition for both functions: blame does not aim to protest against a

claim that is not close to violating any norms at all, and it does not aim to communicate about a conduct that did not happen to begin with. But this does not necessarily mean that there is a more fundamental function underlying blame's protest and communicative functions; instead, the implication is that there are background conditions that constrain the success conditions of blame's functions.

I thus think that the earlier argument does not support the protest function as the more significant function. Another argument is that the communicative function is in fact more significant, since the protest function loses its point if it does not lead to communicating the content of one's protest with one's interlocutors. My response is similar. In the best scenarios, successful protest leads to successful communication. But many real-life scenarios are far less ideal. Moral protest still has a point even when it does not lead to any successful communication. It still has some *extrinsic* positive values that I have identified earlier in §4.4, such as promoting self-respect and reaffirming one's moral beliefs. But successful protest also has the *intrinsic* value of getting things right, simply by one's registering a truth and challenging the aspects of a wrongful conduct that should be challenged.

My tentative conclusion is that neither of blame's functions is, as a general matter, more significant than the other. They reflect two different moral ideals that are equally or similarly important. To decide which function we ought to prioritize in Blame Conflict, we would have to examine the priority between the two functions in a specific context, which I will turn to next.

4.6.2   Context-Specific Significance and Higher-Order Functions

Another question we can ask is not about the general primacy between blame's functions but about which function is *more significant in the relevant kind of contexts*. As an analogy, a ring both has a function of indicating one's marital status and a function to decorate. There may not be much that we can say about which of the two functions is generally more significant. But it is clear that the former function is more significant when it comes to the context of an anniversary

dinner, and the latter is more significant when it comes to the context of a fashion show. Similarly, part of the answer about what to do or feel in Blame Conflict depends on which is the *more significant function in a given context*.

One kind of cases where the significance of blame's functions may vary is when we have different sorts of *social relationship* with the wrongdoer. For example, we may contrast blaming a close friend, a romantic partner, or a close family member on the one hand, and blaming a total stranger on the other. There has been some empirical evidence suggesting that blame may have different sorts of focus in these two kinds of contexts. Arriaga and Rusbult (1998) contrast "partner perspective taking" with "general perspective taking." Partner perspective taking refers to the tendency to "adopt a close partner's point of view in reacting to a given interdependence situation, attempting to feel and think as the partner would feel and think," whereas general perspective taking refers to the general tendency to adopt someone's point of view in this way (Arriaga & Rusbult 1998, 944). They find that participants' different tendencies of general perspective taking do not strongly correlate with their levels of blame; however, participants who are better at partner perspective taking are inclined to blame their partners less. To the extent that perspective taking seems especially important for blame's forward-looking purpose involving communication, Arriaga and Rusbult's studies provide some support for the claim that this forward-looking function may play a distinctively significant role when it comes to blaming one's partners. Aside the empirical evidence, the claim has some intuitive appeal: blame seems governed *to a greater degree* by the function of promoting and facilitating communication when it comes to blaming close friends or romantic partners than blaming total strangers. This provides an example of how a function of blame may be especially significant given a certain kind of context.

Another kind of factor that the pluralist function-based analysis may illuminate concerns whether our response to a Blame Conflict would fulfill some important *higher-order function*. The thought is that, even if neither of blame's functions is obviously more valuable than the

other, our decision to *prioritize* one over the other may itself fulfill some significant higher-order goal. The general idea can again be illustrated using *Unjust System*. I have suggested that this case indicates a global danger for people like Yue to lose a powerful moral tool—blame with its efficient and non-alienating agency-shaping power. Blame's normal operating is disrupted. One way to think about the issue of what Yue ought to do is to ask what *second-best* tool Yue has, and that should be what she resorts to. But another way to think about the issue is to ask what Yue can do that may help *restore* the normal functioning of the blame system. And my proposal is that Yue's blame can be part of what leads to this restoration. This is because Yue's blame, in this sort of conflicts, can express her decision to prioritize one of blame's functions over the other. By insisting blame, Yue not only protests against a first-order wrong done to her, but also protests against, and signals her dissatisfaction of, a higher-order wrong done to her—the wrong that causes the function disruptors to appear, and makes Yue lose the powerful tool of efficient, non-alienating blame, in the first place.

Cummins's system theory of functions is especially illuminating when it comes to the second-order function of blame. Remember that, according to the system theory, blame's first-order function is determined by the dispositions that blame has that can contribute to the proper functioning of a broader system. I have suggested that this broader system involves our moral practices in general. Blame's second-order function, then, will be determined by the dispositions that blame has, in a certain kind of contexts where blame's first-order functions conflict with each other, that can contribute to the same broad system of moral practices. This invites us to examine if prioritizing one of blame's first-order function over the other is something that can by itself contribute to what blame's first-order functions are meant to contribute to in the first place. We can similarly define the second-order function of the practice of *refraining from blame*, in terms of what refraining from blame in a particular context can, by prioritizing one of blame's functions over the other, contribute to the capacities of the system of moral practices.[11]

---

[11]I have focused on the second-order functions, but there is nothing that prevents my account from also asking questions about the prioritization of which second-order function would fulfill an important third-order function, etc.

Given this way of understanding second-order functions, we can see more clearly why Yue's blame would fulfill a significant second-order function. It would express a prioritization of blame's backward-looking function over its forward-looking function; this prioritization is specific to the relevant kind of context where blame regarding systematic social injustice is counterproductive, partly because of the non-ideal aspects of social reality. What can this context-specific prioritization accomplish? First, it registers and protests against a higher-order fact about what Srinivasan calls the "false dichotomy between anger and reason" (2018, 142) in our social reality. The thought is that part of why Yue's blame is unlikely to fulfill its communicative function is because people in Yue's society tend to have a false but common belief: someone who expresses angry emotions cannot be at the same time a good reasoner. They assume a false dichotomy between passion and reason. This assumption is a higher-order fact because it describes the very social conditions in which blame is experienced, expressed, and conveyed. It is also a higher-order fact that is worth registering, marking out, and challenging. Second, the prioritization of blame's backward-looking function is in itself an effort to communicate with the moral community regarding this false dichotomy, a call to improve this non-ideal aspect of our community, and a signal that the victim herself is committed to taking part in this collective effort. In a nutshell: by insisting on blame in *Unjust System*, Yue registers and calls attention to the non-ideal social contexts that make for the source of the first-order conflict between blame's two functions. This registration and calling will make a positive contribution to the proper functioning of the broad system of moral practices in general, and it constitutes an important second-order function of blame.

One possible objection is that this appeal to second-order functions mischaracterizes our phenomenology. Yue's blame, for example, seems to have changed from being about first-order injustice to being about higher-order injustice, and this may seem to come apart from what people like Yue typically experience as the content of their blame. My response here is similar to my

---

This regress line of inquiry, however, will cease being intelligible or stop producing useful results at some stage.

response to the wrong-kind-of-reason objection in §4.5.2: Yue can *adopt the relevant higher-order function of blame as one of her ends*, and there is nothing irrational or psychologically unrealistic for her to do so. Indeed, I think that it offers a fairly natural characterization of people's phenomenon in this kind of double-bind situations: our anger and blame quite commonly escalate from being about a first-order wrong (e.g., racist behaviors) to about a higher-order wrong, about why we are in this kind of troublesome situations to begin with (e.g., social injustice that makes for the source of the relevant kind of double-bind situations).

4.7 Concluding Remarks

I have motivated and defended a pluralist function-based framework that understands many cases of Blame Conflict in terms of conflicts between blame's two functions. I have also discussed various kinds of factors that we should consider in order to resolve a Blame Conflict. But there may still be cases left that are irresolvable. That is, it may be that some instances of Blame Conflict involve conflicting functions that are equally significant even when we examine the particular context, and blame would not fulfill any significant higher-order function in the context either. In such cases, perhaps the most we can provide is a *disjunctive guidance*: that one ought to either blame or not blame.[12] But I believe that the pluralist function-based analysis has more to provide for many irresolvable Blame Conflicts as well. In particular, it can inform us how we and others can *interpret or construe what we are doing* when making either of the two options. Our blame or refraining from blame may carry its distinct interpretive significance due to the function of blame that would be realized. This, however, would require further investigation.

There may be interesting connections between my account of Blame Conflict and the broader debate about moral conflicts in general. One especially difficult question in the broader literature on moral conflicts and moral dilemmas is what justifies the fact that a moral requirement

---

[12]See Brink (1994, 241) for a disjunctive analysis of irresolvable moral conflicts in general.

can remain standing even when it contradicts the all-things-considered action-guidance (see, e.g., Gowans 1996; Tessman 2014). What explains the fact that we still feel regret, guilt, or remorse about failing to satisfy such a moral requirement? If something like a function-based analysis can be extended to some moral conflicts in general, then we may be able to explain this fact by claiming that one's action fails to achieve one of the many functions of the sort of practice that one is engaged in. It is unclear how widely the analysis could be extended, but it is at least very promising for practices closely related to blame, such as praise and forgiveness. It is another line of interesting future work to explore if there are similar conflicts in those practices, and if we could understand them in terms of a pluralist function-based analysis as well.

Chapter 4, in part, is currently being prepared for submission for publication of the material (as "Blame Conflict," by Wang, Shawn Tinghao). The dissertation/thesis author was the primary investigator and author of this paper.

Chapter 5

Response-Dependence in Moral Responsibility: A Granularity Challenge

5.1   Introduction

Reactive attitudes apparently play a crucial role in our practices regarding moral responsibility. When we see a person responsible for a wrongdoing, our natural reaction is often to feel resentment or indignation; when we ourselves do something wrong, we naturally tend to feel guilt about our own actions. What exactly, then, is the relation between these reactive attitudes like resentment, indignation, and guilt and moral responsibility? The following biconditional is widely accepted: a person is morally responsible for *X* if and only if it is appropriate to hold reactive attitudes toward her with respect to *X*. However, when it comes to the underlying metaphysical relations, there is a deep disagreement. Response-independence theorists think a person is an appropriate target for reactive attitudes in virtue of being morally responsible; by contrast, response-dependence theorists think a person is morally responsible in virtue of being an appropriate target for reactive attitudes. Recently, the debate has received much attention. Those who are sympathetic to response-dependence suggest that it offers a simple and unified

account of the conditions under which people are responsible (Shoemaker 2017; Carlsson 2017; also see Todd 2016), it characterizes responsibility in ways analogous to concepts like humor (Shoemaker 2017), it explains the centrality of social community to moral responsibility (Watson 2014; Menges 2017a), and it promises a better answer than its alternatives to the question of how people can be responsible even if metaphysical determinism is true (see, e.g., Strawson 1962; Watson 2004; Beglin 2018).[1] Response-independence theorists, by contrast, are resistant to granting reactive attitudes a fundamental status, suggesting that they instead gain their normative significance by tracking independent facts about people's being responsible (e.g., Fischer & Ravizza 1998; Smith 2007; Zimmerman 2010; Brink & Nelkin 2013; Brink 2021).

However, both sides in the debate have focused almost exclusively on the explanatory relation between reactive attitudes and the binary notion of *being morally responsible*.[2] But it is also intuitive that a person can *be morally responsible to a greater or lesser degree*. Given this, an approach like response-dependence ought to be subject to a test of gradability: can response-dependence be applied to this scalar notion of moral responsibility? In particular, do reactive attitudes have the level of granularity to ground facts about degrees of moral responsibility? These questions have not been addressed in the literature.

This chapter aims to apply such a gradability test to response-dependence in responsibility. I argue that doing so raises a new puzzle for the approach: even if response-dependence is attractive at a comparatively coarse-grained level, it turns out to be too coarse-grained to ground facts about *degrees* of moral responsibility. Though there are several available strategies one may attempt to address the worry, they all come with substantial costs and are difficult to defend. In general, the suggestion is that we should reevaluate, and properly restrict, the metaphysical role of reactive attitudes in theorizing of moral responsibility.

My discussion will focus primarily on *negative* moral responsibility, that is, moral respon-

---

[1]Note that it is not always clear the extent to which the above authors endorse response-dependence. However, they all clearly identify the view as occupying an important conceptual space.

[2]For an exception, see Brink (2021). Also see Nelkin (2016b) on degrees of responsibility.

sibility for things that are morally wrong (or morally bad), rather than in cases of *positive* moral responsibility, that is, moral responsibility for things that are morally right (or morally good). In particular, I will remain agnostic with regard to whether there is a similar granularity challenge against the connection between reactive attitudes and positive moral responsibility.

5.2   Response-Dependence in Moral Responsibility

More accurately, the *response-dependence view* can be put as follows:

(I)   A person is morally responsible for something morally wrong (or morally bad) if and only if, and in virtue of the fact that, it is appropriate to hold some negative reactive attitude toward this person with respect to that thing.[3]

The *response-independence view*, by contrast, states that the reactive attitude is appropriate if and only if, and in virtue of the fact that, the person is morally responsible. Both sides in the debate agree that the propriety condition for reactive attitudes is necessary and sufficient for ascribing responsibility; their disagreement concerns whether the order of explanation is left-to-right or right-to-left.

Response-dependence theorists disagree on what the relevant negative reactive attitudes are. The most common candidates are resentment, indignation, and guilt. But Shoemaker (2017), a defender of response-dependence, suggests that it is anger rather than resentment that figures in (I). His main worry is that "resentment" has been sometimes used in a way that implies a cognitive element of the attitude: a judgment that someone is morally responsible for a wrongdoing. But this cognitive construal of resentment almost immediately contradicts response-dependence. After all, the very idea of response-dependence is to ground facts about being responsible in facts about reactive attitudes. If reactive attitudes themselves are to be explained by their propositional contents, which further involve facts about being responsible,

---

[3]Some versions of response-dependence might prefer replacing "appropriate" with "fitting," "reasonable," "apt," etc. I use "appropriate" as a broad normative notion that incorporates all those formulations.

then the explanation from reactive attitudes would be ultimately circular—which is not a desirable feature for theories of responsibility.[4] Still, as Shoemaker acknowledges, we can also understand resentment in a non-cognitive manner such that it does not necessarily involve any particular cognitive element (2017, 494). I will proceed to use "resentment" in this latter way, so that treating resentment as the paradigm attitude in (I) does not beg the question against response-dependence. I think this ultimately captures the same emotion as intended by Shoemaker's term "anger."[5]

Another issue about (I) concerns what it means for a reactive attitude to be *appropriate*. Important for the response-dependence theorist, the reactive attitudes are appropriate not because people are morally responsible. Rather, what grounds the propriety condition is to be found in the very facts regarding, or *internal* to the practices involving, how we hold reactive attitudes toward each other. The grounding relation can take a variety of different paths. However, if response-dependence were to have any initial appeal, it cannot characterize the propriety conditions in very local terms, for example, according to whether the wrongdoer is actually held to account. It must allow for local fallibility, that is, cases in which a person is morally responsible even though no one actually holds reactive attitudes toward her. This means that the response-dependence theorist should appeal to certain global rather than local features in our responsibility practices:

> (II) It is appropriate to hold some reactive attitude toward a person for something *if and only if, and in virtue of the fact that*, doing so would accord with certain global features in our practices of holding reactive attitudes toward people.

This characterization is general enough to leave open what exactly the relevant global features are. One way to proceed is to examine the simple *dispositional* features, such that the reactive attitude is appropriate if and only if we are disposed to have the reactive attitude under some *standard condition*. Though this kind of approach is sometimes mentioned and briefly

---

[4]That is, unless one endorses a mutual dependence view according to which facts about being responsible and facts about reactive attitudes are mutually dependent. McKenna (2012) suggests such a view. I will put this view aside and instead focus on response-dependence and response-independence, both of which take the order to dependence to be one-way rather than mutual.

[5]Carlsson (2017) argues that guilt, instead of resentment or indignation, is the reactive attitude that figures in (I). Though I will focus more on resentment, I believe that similar issues about granularity will arise for the emotion of guilt as well.

discussed in the moral responsibility literature (e.g., Todd 2016; Shoemaker 2017), it is more fully discussed in the literature on response-dependent conceptions of value. In this broader literature, the "standard condition" can be either just a condition that we normally satisfy (that is, without dispositional masks, finks, etc.) or put in more idealized terms. For example, Lewis takes it to be a condition in which a person gains "the fullest imaginative acquaintance that is humanly possible" (1989, 121); Railton takes it to be a state in which a person can "contemplate his present situation from a standpoint fully and vividly informed about himself and his circumstances, and entirely free of cognitive error or lapses of instrumental rationality" (1986, 16). Both Lewis and Railton defend response-dependence theories of value in general, rather than of responsibility in particular; however, one can easily apply what they say about the standard or ideal condition for dispositions to cases involving moral responsibility: reactive attitude is appropriate if and only if we are disposed to have the reactive attitude once we "gain the fullest imaginative acquaintance that is humanly possible," or once we contemplate our present situation from a standpoint "fully and vividly informed" about ourselves and our circumstances, and "entirely free of cognitive error or lapses of instrumental rationality." The other way to proceed is to examine whether the reactive attitude would be *fitting*. Shoemaker (2017) adopts this approach. Fittingness is about correct representations. Just as fear is fitting when it tracks a dangerous object, according to Shoemaker, moral anger is fitting when it tracks a property that triggers our "anger sensibilities" and belongs to "the sort of properties to which we humans are built to respond with a heated demand for acknowledgment or a tendency to retaliate" (Shoemaker 2017, 510). In contrast to the simple dispositional approach, Shoemaker's theory finds the relevant features in the dispositions of those refined and developed individuals who have a good anger sensibility (2017, 511).[6]

---

[6]Another influential account that is often referred to as being response-dependent is Wallace's (1994) fairness view, according to which reactive attitudes are appropriate just in case they are *fair*. However, as Todd (2016) has pointed out, there are important differences between Wallace's view and other theorists who endorse or are sympathetic to response-dependence. In particular, Todd makes the interesting observation that Wallace never identifies traditional accounts of responsibility (e.g., libertarianism) as taking the wrong side in the order of explanation, in contrast to other response-dependence theorists. I'm inclined to think Wallace's view is open to both response-dependent and response-independent interpretations, depending on how we understand the notion of fairness. Wallace is therefore the target of the current chapter only insofar as his notion of fairness can be explained in a purely response-dependent

(I) and (II) combined together entail the following thesis:

> (III) A person is morally responsible for something morally wrong (or morally bad) *if and only if, and in virtue of the fact that*, holding some negative reactive attitude toward this person with respect to that thing would accord with certain global features in our practices of holding reactive attitudes toward people.[7]

(III), I think, is broad enough to incorporate the various strategies that one can adopt to articulate the thesis of response-dependence.

5.3   The Granularity of Reactive Attitudes

Response-dependence is typically presented, defended, and criticized as being concerned with a binary notion of moral responsibility, as the thesis that whether one *is morally responsible or not* is to be explained in terms of whether one is an appropriate target for negative reactive attitudes such as resentment, indignation, guilt, or anger. But it is intuitive that moral responsibility comes into degrees: one can be *more or less morally responsible* for one's conduct. Given this, it seems that we should also ask whether response-dependence can be applied to this scalar notion of moral responsibility. However, this connection has not been addressed in the literature.

One may think the application is just a natural extension of (III): as we can ground facts about binary responsibility in the appropriate reactive attitudes, we can ground facts about scalar responsibility in the *appropriate degrees* of reactive attitudes. However, this then raises new problems and challenges regarding the approach. In this section, I contend that reactive attitudes, such as resentment and indignation, are fairly coarse-grained in terms of their degrees; at least, this is so according to how response-dependence theorists characterize the nature of such attitudes. I will contrast this with the granularity of facts about degrees of moral responsibility in the next section.

---

way.

[7]The inference relies on the widely accepted assumption that grounding is a transitive relation.

To start with, what does it even mean to claim that emotions like resentment have degrees? On perhaps the most natural understanding, the claim just means that reactive attitudes have different degrees of *intensity*. There is no denying that emotions have varying intensities. Some differences in intensity implicate slightly different attitudes. For instance, we might distinguish between sadness (mild) and sorrow (strong). There can also be different levels of intensity within a given reactive attitude. For instance, we can feel resentment to different degrees, measured by a scale from mere annoyance or frustration, to relatively stronger forms of anger, and finally to rage. However, the scale regarding degrees of emotions is usually very rough in its measure. Take happiness as an example. Suppose that Daniel feels happy about getting a monetary reward for teaching excellence. Depending on how much money the reward involves, Daniel can feel more or less happy—perhaps something like "a little bit happy" around $5, "somewhat happy" around $50, and "very happy" around $500. But I take it that Daniel does not feel happier when it is $57 instead of $56. I stress that this is not just because Daniel cannot intentionally adjust degrees of happiness, or because Daniel does not care about the value of 1 dollar—rather, it is because our emotional states are carved out in human psychology in a way such that they do not have the level of granularity required to register only slight differences in values. Now, there is surely a sense in which one can maintain that Daniel does feel happier about $57 than about $56, if by claiming so one already implies that Daniel *believes* $57 is a little bit more valuable. However, putting aside this cognitive element that arguably may come apart with the emotion itself, degrees of emotions usually seem fairly coarse-grained.

The failure to register small differences can be applied to negative reactive attitudes as well. Imagine Heru, a somewhat mentally impaired person, who stole Ann's passport. Now compare Heru with Henry, who stole Ann's passport at a different time, but is just *slightly less mentally impaired* than Heru. I take it that Ann does not resent Henry more than Heru, even though Henry is slightly more competent. Again, this is not just because Ann cannot intentionally adjust degrees of resentment. Instead, it is because our emotional states simply do not have the

100

level of granularity required to register only a slight difference in moral competence. That is just the way in which emotions are carved out in human psychology. I will refer to the general insight here as the *thesis of limited information registration*: human emotions are set up in such a way that they only track potentially responsibility-relevant factors, such as moral competence, in a fairly coarse-grained manner. More specifically, one way of measuring Heru and Henry's relevant mental/moral competence is to look at the proportion of relevant possible worlds in which they could have acted according to moral reasons and thereby refrained from stealing Ann's passport or committing similar wrongdoings. Suppose this percentage is 40 percent for Heru but 41 percent for Henry. It seems unlikely Ann's emotion would change according to this slight increase in the proportion of possible worlds. I stress that the point here does not hinge on the possible-world account of moral competence. It would go through on other accounts as well, as long as moral competence is characterized in a fairly fine-grained manner. The general thought is that there are many cases in which, just like Heru and Henry, our emotions do not have the level of granularity required to differentiate between very similar but slightly different moral agents.[8][9]

Again, there is surely a sense in which we may say that Ann resents Henry more than Heru, if this already entails, or even just means, that Ann believes that Henry is a little bit more morally responsible than Heru for the conduct. But it is unclear whether this sort of cognitive judgment is part of the emotion of resentment. More importantly, response-dependence theorists would certainly deny that it is. As I have suggested in §5.2, the relevant reactive attitudes that figure in the response-dependence approach should not involve a particular cognitive element, and surely should not involve any cognitive judgment about moral responsibility, in order to avoid circular explanation. But if we put aside this cognitive element, then the thesis of limited

---

[8]There are parallel cases where two agents have the same capacities, but one's action is slightly more wrong than the other's. Imagine that Fei stole $100 from Ann, whereas Fara stole $101 from Ann. Though Fara's action is slightly more wrong than Fei's, and Fara is thereby more blameworthy, Ann's resentment does not seem to have the level of granularity required to track this difference. To the extent that response-dependence theorists about moral responsibility also endorse response-dependence about blameworthiness, cases like this can be used to construct a structurally similar granularity challenge for them as well.

[9]For a related discussion about proportionate justice in the criminal law, see Brink (2021, 390–2).

information registration becomes highly plausible.

To further elaborate the thesis, it is useful to consider an analogy to degrees of belief. The orthodox Bayesian view that beliefs have precise credences is frequently criticized as psychologically unrealistic (see, e.g., Mahtani 2020). It is very rare (unless perhaps in certain cases involving precise mathematical questions in particular) for one to actually have 0.64 credence in any proposition. Note that this is still consistent with thinking beliefs are gradational in some sense. Our beliefs do register information about changes in probabilities, but only do so in a fairly coarse-grained manner. They do not, and do not have the level of granularity required to, register only slight differences in probabilities. My contention is that a similar story is true of degrees of reactive attitudes—they do not have the level of granularity required to register small differences in certain potentially normatively relevant features of moral agents, like their moral competence. It is psychologically unrealistic to posit the existence of fairly precise reactive attitudes.

From the evolutionary perspective, it is hardly surprising that information registration in emotions is limited and selective. Emotions have many important roles to play, but one of them is to trigger immediate actions when receiving certain feedback from the outside world. The evolutional pressure is likely to associate our emotions with something that matters, but it is unlikely to make the association very accurate; coding every aspect of the external world that matters would be too costly to facilitate the efficient emotional responses that conduce to our survival.

My claim is that emotions are limited and selective in information registration, which is compatible with their registering some information in useful ways. This fits well with the current empirical evidence we have. One common measure of emotional intensity in empirical psychology is by appeal to self-report, and the measure typically adopts a coarse-grained scale, instead of, say, offering a twenty-point scale with low-to-mid-moderate, etc. For example, Larsen et al. ask the participants "how happy are you?" using a rough scale ranging from 1 to 5, where 1,

3, and 5 are labeled as "slightly," "moderately," and "extremely" (2001, 688). Further, I think some self-report data constitute indirect evidence for the coarse-grainedness of emotional states. One potential source of evidence comes from Barrett's (2006) work on individual differences in people's ability to describe their emotional experiences in precise terms—an ability that she calls "emotional granularity." She distinguishes between "low granularity" and "high granularity" individuals. The low granularity individuals describe their emotional states in very general terms. One example goes like this: "I just felt bad on September 11th, really bad" (Barrett 2006, 38). By contrast, the high granularity individuals describe their emotions using more precise labels. One example, also in response to 9/11, goes as follows: "My first reaction was terrible sadness... But the second reaction was that of anger, because you can't do anything with the sadness" (Barrett 2006, 38). The lesson here is not only that the low granularity individuals lend support to the coarse-grainedness of emotions. Even the high granularity individuals differentiate negative emotions using relatively general labels such as "terrible sadness" and "anger"; they do not describe their emotional states using more precise degrees. I admit this does not directly show that people's emotional states are coarse-grained; self-report can be limited in various ways, and people might lack concepts to describe a genuine phenomenon. Still, self-report is considered a major way of measuring emotions, and, therefore, the findings in emotional granularity give some indirect support to the claim that emotions are coarse-grained. As any empirical thesis, my claim here is subject to further scrutiny; however, given the evidence we have thus far, I take it that our default thesis should still be that reactive attitudes are fairly coarse-grained.

5.4   The Granularity of Responsibility Facts

In this section, I will argue that, in contrast to reactive attitudes, facts about moral responsibility have fairly fine-grained degrees. I stress that this does not mean there are always *precise cardinal numbers* assigned to degrees of responsibility; I agree that facts about responsibility are

not that fine-grained. Rather, the contention is that they are still fairly fine-grained, and, most important for my purpose, more fine-grained than our reactive attitudes are able to capture.

My argument has two steps: first is to argue that there is a scalar notion of moral responsibility; second is to argue that, if we take for granted that there is a scalar notion of moral responsibility, then there are good reasons to believe moral responsibility comes in fairly fine-grained degrees. The first step, I believe, is quite straightforward, since a scalar concept of responsibility is necessary for the purpose of arriving at a proper assessment of one's degrees of blameworthiness. It is plausible that one is morally blameworthy just in case one is morally responsible for something morally wrong (or morally bad). And surely one can be morally blameworthy to different degrees, since one can deserve more or less blame. Further, it also appears one can deserve more or less blame even if we hold fixed the degree to which one's action is morally wrong (or morally bad). But this scalarity would be mysterious if moral responsibility is simply a binary notion. In reply, one might point out that there have been important theories of moral responsibility that take it to be an all-or-nothing, threshold notion (e.g., Fischer & Ravizza 1998; Fischer 2004). However, the fact that there is a threshold concept of responsibility does not entail that moral responsibility does not come in degrees. Rather, the more plausible view is that there are both a threshold and a scalar concept of moral responsibility, just as in epistemology, where many agree that there are both a threshold notion of belief and a scalar notion of degrees of belief. It is thus unsurprising that even Fischer and Ravizza's (1998) threshold framework has been later extended to applications involving degrees of responsibility (see, e.g., Coates & Swenson 2013), since the most plausible normative theory should provide both the threshold and the scalar notion.[10]

Now I turn to the second step of the argument: if we take for granted that there is a scalar notion of moral responsibility, then there are good reasons to believe moral responsibility comes in fairly fine-grained degrees—at least more fine-grained than our reactive attitudes are able to

---

[10]Also see Nelkin (2016b) and Brink (2018, 2021) for, broadly speaking, reasons-responsiveness approaches on degrees of responsiblity.

capture. The reason for this is that the opposing view, according to which moral responsibility comes in degrees but is no more fine-grained than the gradability of our reactive attitudes, is not attractive, once we consider how such a view is related to substantial conditions of moral responsibility. While some response-dependence theorists, such as Shoemaker, may not believe that any *necessary or sufficient* conditions of responsibility can be offered without appealing to the reactive attitudes themselves, they also do not think it is irrelevant what *substantial conditions* end up being captured by the response-dependence theory. Indeed, I think one important reason why any response-dependence theory is attractive is because it seems to fit pretty well with the *plausible substantial conditions of moral responsibility*. Here is what I mean. It is generally intuitive that certain substantial conditions—again, note that they do not have to be necessary or sufficient conditions—can affect degrees of responsibility, including, for example, the reasons-responsiveness condition and the quality of will condition. The reasons-responsiveness condition specifies whether and to what degree a moral agent or her mechanism has the ability to respond to reasons;[11] and the quality of will condition specifies whether and to what degree a moral agent's action manifests her good or bad quality of will.[12] It is intuitive that such factors can affect degrees of moral responsibility, and it is also true that our emotion, at a coarse-grained level, does vary according to those conditions. After all, reactive attitudes come with *some* degrees. Thus at a coarse-grained level, we do resent a mature, fully reasons-responsive agent more than a mentally impaired, less than fully reasons-responsive agent, when other things are equal; similarly, we do resent a person more for an action that manifests a really bad quality of will than for an action that manifests a somewhat bad quality of will, when other things are equal. If response-dependence were to not arrive at those results even at the coarse-grained level, then, I think, the view would not be attractive at all, since it asks for too much conceptual revision. That is, the fact that response-dependence seems to capture pretty well the plausible substantial

---

[11]See Fischer & Ravizza (1998), Fischer (2004), Wolf (1990), Nelkin (2011), Vargas (2013), McKenna (2016b).
[12]See Scanlon (2008), Smith (2008), Talbert (2008), Hieronymi (2004), McKenna (2012, 2016a), Shoemaker (2015).

conditions of moral responsibility is a central attraction for the view. This can also be seen by Shoemaker's defense of response-dependence, where he contends that the approach explains how features like what I have referred to as the plausible substantial conditions of moral responsibility are all "blameworthy-makers" and matter for responsibility in a way consistent with the "apt variations in type and degree of anger" (2017, 510).

My contention is that this attraction is, unfortunately, seriously undermined. This is because a closer examination shows that response-dependence *only* gets the intuitive results about the plausible substantial conditions at the coarse-grained level, and once we examine cases about, for example, two agents with only slightly different reasons-responsiveness abilities or slightly different qualities of will, response-dependence no longer amounts to different normative judgments for those agents. My earlier case about Heru and Henry illustrates this for the reasons-responsiveness condition: though Heru is slightly less capable of responding to reasons than Henry is, this is too small a difference to be reflected in the patterns of our reactive attitudes. And this undermines the attraction of the response-dependence view not because the two agents should be morally responsible to different degrees from a response-independent perspective (which would beg the question); rather, the reason is that they should be morally responsible to different degrees from the perspective of the plausible substantial conditions. We find the reasons-responsiveness condition intuitive not just because their effects only exist at the coarse-grained level. Instead, as illustrated in the case of Heru and Henry, our intuition supports that reasons-responsiveness should affect degrees of responsibility both at the coarse-grained and the fine-grained level. Thus, in the absence of a plausible debunking story, the apparent attraction of response-dependence is seriously undermined once it is shown that the view cannot really capture the full range of applications of our general intuition.[13]

The same form of argument can be made about the quality of will condition. Theorists differ on how to cash out the notion of quality of will, but it is often taken to be a sort of care or

---

[13]See Coates & Swenson (2013), Nelkin (2016b), and Brink (2018, 2021) for alternative, but also fairly fine-grained, characterizations of degrees of responsibility in a reasons-responsiveness framework.

regard that one has of other people's moral interests. Care or regard may be similar to reactive attitudes in terms of their granularity. But the degree of *manifestation* of a bad quality of will can still be fairly fine-grained. To see this, note that manifestation typically implies a causal or explanatory relation. But the substandard quality of will does not have to be the full causal or explanatory story; it can just be a partial cause or partial explanation. At the intuitive level, the concepts of "more of a cause" and "more of an explanation" can be fairly fine-grained. Recent work on degrees of causation in metaphysics (for a review, see Kaiserman 2018) adds further support to this intuitive claim by laying out the various models in which the degree of causation can be fine-tuned. We can then construct cases similar to that of Heru and Henry. Imagine that Dane has a substandard quality of will toward her roommate, and this partially causes or explains her stealing her roommate's yogurt from the fridge. But this is not the full causal or explanatory story, since Dane is also under some pressure from another person, Daniel, to steal the yogurt. We can then vary the extent of the causal or explanatory role that Dane's bad quality of will (rather than the pressure from Daniel) plays, in a way that the difference can be too small to be registered in the resentment of Dane's roommate. That is, though response-dependence initially appears attractive in virtue of being able to capture the quality of will condition, it turns out that our intuition supports that the quality of will condition should affect degrees of responsibility both at the coarse-grained and the fine-grained level. Thus, response-dependence does not really capture the full range of applications of the quality of will condition, and the initial attraction of response-dependence is again seriously undermined.

Here is another way to put the dialectic so far. Response-dependence and response-independence are *meta-normative* theories. They are committed to a particular order of explanation, but do not by themselves offer substantial conditions for being responsible. By contrast, there are *substantial* conditions like reasons-responsiveness and quality of will. A theoretical virtue for a good meta-normative theory is that it can be combined with the most plausible substantial conditions as well as their intuitive applications. Since response-dependence can only

be combined with a very narrow kind of applications—applications at a coarse-grained level—of the substantial conditions, this shows that response-dependence, as a meta-normative theory, has a serious theoretical cost. This cost is obviously worrisome for a neutral party who has not yet committed to response-dependence or response-independence, but I believe that even someone already committed to response-dependence also faces a serious challenge here, given that one major attraction of response-dependence—that it appears to accurately capture the substantial conditions of responsibility—is seriously undermined.

To sum up, my contention is that there is a mismatch challenge against response-dependence:

(1) Facts about the reactive attitudes are fairly coarse-grained.
(2) Responsibility facts are fairly fine-grained.
(3) Both (1) and (2) are true when the standard of granularity is held fixed.
(4) Thus, responsibility facts cannot be fully determined by facts about the reactive attitudes.

Let me emphasize that coarse-grainedness and fine-grainedness are relative terms; central to the challenge is the comparative mismatch between the two measures of granularity, which is why (3) is a necessary step of the inference. I have defended (1) in the last section and (2) and (3) in the current section. Combined, the conclusion (4) seems to contradict the response-dependence view, which attempts to grounds responsibility facts in facts about reactive attitudes.[14]

---

[14]There is a different perspective to look at the mismatch in degrees of reactive attitudes, by appealing to the acceptable stability and variation for appropriate reactive attitudes. It seems the propriety condition for reactive attitudes should allow for a reasonable degree of instability. It is typically not inappropriate for a person to feel a certain degree of resentment toward a wrongdoer at one time, but at a later time feel a bit less resentment toward the same wrongdoer. This can be true even when the person neither forgives the wrongdoer nor receives any further information. If both the initial, heated-up emotion and the later, cooled-down emotion are appropriate, how can the response-dependence theorist tell a story of grounding degrees of responsibility using degrees of emotions? More work needs to be done to tell a convincing story here. Bykvist (2009) appeals to similar concerns to argue against fitting attitude theories of value. He contends that one difficulty consists in the fact that "how strongly we should react emotionally seems also to depend on temporal matters" (Bykvist 2009, 16).

5.5 Possible Replies to the Mismatch Challenge

5.5.1 The Idealization Strategy

A response-dependence theorist may reply by appealing to the difference between the actual and the ideal or *appropriate* reactive attitudes. The suggestion is that, even if we in fact do not have fine-grained reactive attitudes, we can still use fine-grained reactive attitudes as idealizations by maintaining that they are the appropriate ones to have. And since the formulation of response-dependence uses appropriate rather than actual reactive attitudes, one might suggest, there will not be any mismatch at all.

However, the current strategy may not be available to the response-dependence theorist. Response-dependence views of responsibility aim to naturalize the propriety condition for reactive attitudes, by grounding it in the global features of our practices of holding reactive attitudes toward people. It is then not clear how, by doing so, they can make the appropriate reactive attitudes more fine-grained than the actual ones. The specifics will of course depend on how one understands the propriety condition. Remember two common approaches are the simple dispositional approach and the fitting approach. The former route finds the appropriate reactive attitudes in our dispositions under standard conditions. The standard conditions are sometimes understood as just the normal conditions, those that we usually satisfy when there are no dispositional masks, finks, etc. But this does not help with the granularity issue. Putting the standard conditions in more idealized terms does not help either, at least according to the classic accounts. Remember that Lewis takes the condition to be one in which we have "full imaginative acquaintance" with the subject, and Railton takes it to be one in which we are fully informed of the circumstances and fully free of cognitive errors. But neither more acquaintance nor being more informed can help make our reactive attitudes much more fine-grained than they already are. Again, both Lewis and Railton aim to defend response-dependence for value in general; however, it is reasonable to expect some response-dependence theorists of responsibility to appeal to very similar ways of

understanding the standard condition as well.

The fitting route finds the appropriate reactive attitude in the fitting responses. Remember that the notion of fittingness is used to rule out those responses of people with bad "anger sensibilities" (Shoemaker 2017). According to Shoemaker, a good moral sensibility is analogous to the wine sensibility of a sommelier, in contrast to that of a diner (2017, 511). One may then suggest that appealing to this refined sort of sensibilities can help avoid the mismatch challenge. However, I am inclined to think that the challenge can also be applied to those with good anger sensibilities. I tend to think that, in the earlier example, Ann does not feel angrier toward Henry than toward Heru, *even if* Ann has the good anger sensibilities. This is because the limitation of information registration is not due to the poor sense of any individual, but due to a general limitation in human psychology. Again, there is surely a sense in which we can develop fairly sophisticated, refined capacities to track moral responsibility facts, if we include among these capacities the cognitive capacity of making judgments about degrees of moral responsibility. However, if response-dependence theorists were to appeal to such a cognitive capacity to understand good anger sensibilities, then the alleged fundamental status of reactive attitudes would again be undermined. But if they were to exclude this cognitive capacity, then the development of anger sensibilities would be greatly limited. It does not seem that we could refine the granularity of anger or resentment that much without employing our capacity to learn and apply what to *believe* about degrees of moral responsibility.

It thus seems that the two common routes of understanding the propriety condition are not helpful for formulating a convincing reply to the mismatch challenge. As a result, for those who want to pursue the idealization strategy in response to the challenge, they would need a new way to explicate the notion of idealization in enough detail and provide independent motivations for favoring it over its alternatives. It is unclear to me how one can do so, but I believe that this will lead to much valuable inquiry.

5.5.2   The Restriction Strategy

On a different kind of reply, one may suggest that response-dependence should be restricted to *facts about whether one is morally responsible*, which is compatible with response-independent explanations for *facts about degrees of moral responsibility*. Indeed, this may even be a reasonable reinterpretation of the actual view of many proponents of response-dependence. The conclusion of my mismatch challenge then, one may suggest, turns out to be consistent with a response-dependence approach about moral responsibility.

I think this is a promising reply, but I worry that it significantly compromises the theoretical payoff of endorsing response-dependence. For example, one motivation for response-dependence is its simplicity; but this would be undermined if we need to supplement the account using response-independent explanations anyway. It is also unclear the extent to which this hybrid picture will help answer the question of how people can be responsible if metaphysical determinism is true. After all, if only facts about whether one is morally responsible is response-dependent, then the approach provides no explanation for why it cannot be the case that everyone is morally responsible to *an extremely small degree* if determinism is true. To say the least, defenders of the current strategy need to explain why a hybrid picture is well-motivated, and, in particular, why it does not already concede too much and still preserves many of the theoretical advantages of the pure response-dependence approach.[15]

5.6   Concluding Remarks

I have raised a granularity challenge against the response-dependence view of moral responsibility. Response-independence theorists, by contrast, do not face the same challenge, since they are free to use factors external to our practices of holding each other responsible in

---

[15]Another way to restrict the scope of response-dependence is to claim that (i) our responses track responsibility-relevant factors such as reasons-responsiveness, and that (ii) this fact explains why factors like reasons-responsiveness are relevant to responsibility. Thanks to David Brink and Craig Agule for suggesting this possibility.

grounding the propriety conditions for reactive attitudes. Thus, the granularity challenge lends support to the response-independence theories. However, it is worth noting that the mismatch in granularity also has important methodological upshots even for the response-independence theorist—for example, the mismatch suggests that we will not be able to use fine degrees of emotions as *epistemological* guides to facts about degrees of responsibility. In general, the upshot is that the mismatch challenge should lead us to reevaluate and properly restrict the theoretical role of reactive attitudes.

It remains an open question whether a similar granularity puzzle is as challenging in domains other than moral responsibility. For example, are our pro-attitudes too coarse-grained to be the metaphysical basis of values? Are the psychological features of respect too coarse-grained to be the metaphysical basis for dignity? Are our legal practices too coarse-grained to be the metaphysical basis for legal culpability? These questions are not to be solved here; but note there is also no need to demand the same kind of answer to this cluster of questions—for example, it might turn out that dignity is a fairly coarse-grained notion and can be fully determined by the psychological features of respect, but value is too fine-grained to be fully determined by our pro-attitudes. The bottom line is that there is a real granularity puzzle for response-dependence in moral responsibility, and the plausibility of response-dependence approaches in other domains should be subject to a similar granularity test.

Chapter 6

Shame and the Scope of Moral Accountability

6.1   Introduction

There are a variety of negative emotions a person may experience in the face of a moral infraction: resentment, indignation, guilt, fury, fear, disgust, shame, and so on. Philosophers often categorize these emotions into two groups: some, including at least resentment, indignation, and guilt, are *reactive attitudes*; others, such as fury and fear, are *non-reactive attitudes*. The labels are attached in terms of each emotion's relation to moral responsibility. Reactive attitudes, unlike non-reactive attitudes, target participants of the practices involving holding each other, and holding oneself, morally responsible. The propriety conditions of reactive attitudes, unlike those of non-reactive attitudes, require that their target is in fact morally responsible. Though non-reactive attitudes may serve important functions, according to the standard philosophical view, it is the reactive attitudes that play the most central role in those activities involving holding people morally responsible.

Philosophers debate the scope of reactive attitudes, and it is particularly controversial

where *shame* fits in. Is shame a reactive attitude? Though Strawson's (1962) seminal paper on reactive attitudes endorses a positive answer, the view has been frequently challenged. Wallace (1994) restricts reactive attitudes to only resentment, indignation, and guilt, but excludes shame, because shame is not constitutively linked to "holding oneself to a demand" (241) as guilt is. Darwall (2006) also excludes shame in his second-personal framework, arguing that shame inhibits the kind of "second-personal engagement" (72) that is, according to Darwall, central to the interpersonal functional role of reactive attitudes. The assumption that excludes shame as a reactive attitude also seems implicit in many other theorists' work, given the disproportionate attention they choose to devote to guilt over shame. Just as one example, Clarke (2013, 2016) offers a defense of retributivism based on the idea that responsible wrongdoers deserve to feel guilt, but he very quickly dismisses the relevance of shame, claiming that it is not "a matter of justice" (2016, 128) that one feels shame about failing to live up to certain standards.

Another possible answer goes beyond the simple "yes" or "no," but rather appeals to a distinction between different senses of reactive attitudes according to their connections to different senses of moral responsibility. Watson (1996) has distinguished between *attributability* and *accountability* concepts of responsibility. Here is a rather minimal way to understand the two concepts, when applied to things that are morally bad or morally wrong, namely, things that we are *negatively* morally responsible for: A person is morally responsible for *X* in the attributability sense just in case *X* reveals her evaluative commitments, namely, what she values. Sometimes attributability is defined in terms of whether *X* reflects one's substandard moral character or traits. I am inclined to think the definition in terms of evaluative commitments is more plausible, but I expect the arguments and views in the current chapter to apply to the character-based definition of attributability just as well as it applies to the value-based definition. By contrast, a person is negatively morally responsible for *X* in the accountability sense just in case *X* provides a pro tanto moral reason (or at least a pro tanto reason that is understood within our practices to be

moral) for other people to hold resentment or indignation toward her.[1] It is also commonly held that, if a person is negatively morally responsible for *X* in the accountability sense, then she is *morally blameworthy* for *X*, which further entails that there is a pro tanto moral reason (or at least a pro tanto reason that is understood within our practices to be moral) for other people to *blame* her. This does *not* always give rise to an all-things-considered reason to blame, because there may be contexts where the pro tanto moral reason is overridden, e.g., when the potential blamer lacks the standing to blame. Sometimes accountability is defined in terms of more substantial conditions such as whether the person *deserves* sanction in virtue of *X* or whether *X* violates an interpersonal demand. I will proceed with the more minimal formulation of accountability, and address these more substantial views about accountability in §6.4. On the basis of the distinction between attributability and accountability, Carlsson (2019) and Portmore (2019a, 2019b) defend an intriguing position that shame is a reactive attitude but it is concerned with the attributability rather than accountability sense of moral responsibility; guilt, they suggest, is concerned with accountability instead. Hence, we have three available answers to the question raised in the beginning: *exclusivism*, the view that shame is not a reactive attitude at all; *inclusivism*, the view that shame is a reactive attitude in the sense that concerns moral accountability; and *divisionism*, the view that shame is a reactive attitude but not in the sense that concerns moral accountability.

In this chapter, I aim to advance novel arguments for inclusivism, what has now become the heterodox view in the literature.[2] Throughout the chapter, I will assume a commonly accepted picture of reactive attitudes: their *targets* are the good will, ill will, or indifference of moral agents as participants of those practices involving holding each other, and holding oneself, morally responsible; and their‿ propriety conditions‿ are as follows: it is appropriate to hold a reactive attitude toward a person just in case the person is morally responsible for something. We can

---

[1]This involves some simplification. Some have thought that accountability only entails a *conditional* pro tanto reason to hold the negative reactive attitudes (see Nelkin 2016a).

[2]There are important exceptions. Watson (2014) offers a brief but insightful defense of inclusivism. Shoemaker distinguishes between different senses of responsibility, but claims what he calls "agential shame" in fact "cuts across all categories" of moral responsibility, including attributability, answerability, and accountability (2014, 26).

then distinguish between *attributability-tied reactive attitudes* and *accountability-tied reactive attitudes*. The former kind of attitude is appropriate toward a person just in case the person is morally attributable for something, whereas the latter kind of attitudes is appropriate toward a person just in case the person is morally accountable for something. There has been a debate about what exactly the appropriateness consists in, and whether it means the same relation for attributability-tied and accountability-tied attitudes (e.g., Shoemaker 2017; Carlsson 2019; Portmore 2019a, 2019b). I will not take a stand on this controversy. Rather, my approach is to start from two assumptions that I believe different parties in the debate would agree on. First, the propriety condition for attributability-tied reactive attitudes is the *fitting condition* of the emotion, which further involves having correct representations. Note that this alone does not ensure that we have any moral reasons to hold the emotion, since many ordinary fitting conditions do not entail any moral reasons at all, for example, when a joke is fitting because it is funny. Second, the propriety condition for accountability-tied reactive attitudes should provide *moral reasons*, or at least reasons that are understood within our practices to be *moral*, to hold the attitude. This is compatible with both the claim that it is not a fitting condition at all (Carlsson 2019) and the claim that it is still a fitting condition but gives rise to moral reasons due to the fact that the content of the accountability-tied attitudes is always about desert, a moral-reason-giving notion (Portmore 2019a, 2019b).

Another preliminary point concerns what it exactly means to say that shame is a reactive attitude. Does it mean some particular *tokens* of shame are reactive, or the general emotional *type* is as such? I am inclined to endorse both. There are some particular tokens of shame as accountability-tied reactive attitudes, and, as a further point, those tokens belong to an emotion type that we can properly refer to as an accountability-tied reactive attitude. It is worth noting that, however, we may have to do some conceptual regimentation to single out this type of shame, perhaps as "moral shame" or "agential shame," in order to exclude, for instance, physical shame about one's bodily features—which intuitively falls outside the domain of reactive attitudes. But

another possibility is that shame is a kind of accountability-tied reactive attitude, even though it has some subcategory, such as physical shame, that we rarely have moral reason to hold and is thereby rarely appropriate. This is a choice point, I think, that ultimately depends on how we individuate psychological kinds, a difficult task that I will leave aside here. Rather, my thesis is that either shame or a subcategory of shame is an accountability-tied reactive attitude. If the thesis holds, this would be a success for the inclusivist, since the current defenders of exclusivism or divisionism apparently think neither shame nor a subcategory of shame counts as an accountability-tied reactive attitude.

6.2   Shame and Guilt: Intertwinedness and Differentiation

Here is a first attempt to develop an argument for inclusivism. Note that both exclusivism and divisionism presuppose that shame and guilt are quite distinct. But this appears to be in tension with the phenomenology that shame and guilt are deeply intertwined with each other. On my usage of the term, two emotions are deeply intertwined with each other when, in a wide enough range of cases, they are *overlapped*, *undifferentiated*, and *convertible to each other*. First, overlapping means that, in many cases, the two emotions are both experienced by a person and often come and go together. Second, they are undifferentiated in the sense that, in many cases, a person does not differentiate between the two emotions, and, upon reflection, it can be hard for a person to tell which of the two emotions she feels. Finally, they are convertible to each other such that intense feelings of one emotion sometimes shift over and become the other, and/or the other way around. Shame and guilt satisfy all three conditions, as supported by both reflection on ordinary experiences and empirical evidence. (i) Though shame and guilt can come apart, they significantly overlap in a large number of cases. This is further supported by experiments showing that only very few antecedents uniquely elicit either shame or guilt, whereas a large majority of antecedents elicit both (Keltner & Buswell 1996; Tangney 1992). (ii) We can find it quite

hard to distinguish if our emotion is shame, guilt, or both, at least for many cases. From college students to clinical psychologists, as Tangney and Dearing (2002) observe, people use "shame" and "guilt" interchangeably and perhaps even inconsistently. It is not uncommon at all for people to talk in terms of the phrase "shame and guilt" and discuss their sources and effects together, with little discrimination between the two emotions. (iii) Guilt experiences appear to be able to be converted to shame experiences (e.g., Lewis 1971). These features involving intertwinedness may appear to pose a prima facie challenge against exclusivism and divisionism, thereby supporting inclusivism. That is, since (1) guilt is an accountability-tied reactive attitude, and (2) shame and guilt are deeply intertwined, we should conclude that (3) shame is also an accountability-tied reactive attitude. I will refer to this as the *Intertwinedness Argument* for inclusivism.

Though I will ultimately contend that the intertwinedness argument, when suitably revised and supplemented, can pose a strong case for inclusivism, the inference as stated in the above form is unconvincing and subject to counterexamples. One may think resentment is deeply intertwined with what Pereboom calls "fury," a more primitive kind of anger that we share with "bears and wolves" (2014, 147). But resentment is, whereas fury is not, a reactive attitude. More generally, the inference seems to overlook the distinction that Wallace (1994) makes between *natural* attitudes and *reactive* attitudes. Wallace argues that inclusivists make the mistake of confusing the two and thinking whatever attitudes that are natural to have in interpersonal relationships must be reactive. He then suggests that emotions like shame are natural but non-reactive attitudes. Similarly, Wallace might reply that intertwinedness shows that both emotions are natural, but not that they should be similarly treated when it comes to categorizing emotions as attributability-tied reactive, accountability-tied reactive, or non-reactive.

Nonetheless, I take it that the intertwined argument can still be modified into a forceful, burden-shifting move in support of inclusivism. This is because deep intertwinedness between the two emotions should at least shift the burden to those who reject inclusivism to show how differences between shame and guilt justify treating only the latter as an accountability-tied

reactive attitude. This burden is met in the comparison case of the difference between fury and resentment. In contrast to resentment, fury involves a distinct self-defense mechanism that registers the information that "there's a threat to be violently neutralized" (Pereboom 2014, 147), which plays little role in our responsibility practices. The task falling upon us then is to investigate if there are similar reasons to differentiate shame from guilt in a manner that justifies categorizing them differently.

Just as we can differentiate fury from resentment despite their intertwinedness, there are ways to differentiate shame from guilt too, in terms of a cluster of features including their phenomenology, representational content, and motivational tendency. Here is a rough characterization of the major differences in these three regards. The phenomenology of shame is an intense painful feeling that can be said to involve a wish to hide or disappear, fused with a disruption in thought, a sense of confusion, and sometimes blushing (Lewis 2008, 748); the representational content of shame is a failure to live up to some norm or expectation, where the failure is primarily concerned with one's substandard self (Niedenthal, Tangney, & Gavanski 1994; Tracy & Robins 2006; Tangney, Stuewig, & Mashek 2007; Lewis 2008); the motivational tendency is to dissipate the emotion by "reinterpretation, self-splitting, or forgetting" (Lewis 2008, 748). In contrast, the phenomenology of guilt is a painful feeling, but generally less intense than shame (Tangney 1995); the representational content of guilt is a particular transgression typically in the form of a norm-violating action (Niedenthal et al. 1994; Tangney et al. 2007; Lewis 2008); the motivational tendency is to direct the agent to perform a corrective action to repair the failure (Lewis 2008, 748).

Many details of this characterization are controversial and up for debate, but I take it to be a good starting point to consider those familiar ways in which shame and guilt can come apart. The characterization also fits well with empirical findings. In particular, the self-action contrast is widely accepted in social psychology: shame is more concerned with failure that reflects one's substandard *self*, whereas guilt is more concerned with failure in a *particular action*. Tangney,

Stuewig, & Mashek claim that shame is "a negative evaluation of the global self," whereas the focus of guilt is "a negative evaluation of a specific behavior" (2007, 349). Shame brings the agent to see a shortcoming in herself, to think "if only I weren't ...," while guilt brings the agent to see a failure in her specific behavior, to think "if only I hadn't ..." (Niedenthal et al. 1994). This is "the currently most dominant basis for distinguishing between shame and guilt" (Tangney et al. 2007, 349) in the psychological literature. In philosophy, many draw a broadly similar self-action distinction between shame and guilt as well. For example, Rawls argues shame and guilt have different explanations, using the example of a person who "cheats or gives in to cowardice" and feels both guilt and shame (1971, 391). According to Rawls, this person's guilt is due to the quality of his action, but his shame is due to a diminished sense of "his own worth" (1971, 391).

Let me stress that the self-action contrast does not require the target of shame to be only about the self or the target of guilt to be only about actions. We can feel both shame and guilt toward an action, but, the self-action contrast suggests, shame about the action is more focused on the ways in which the action reflects on the person's self, whereas guilt about the action focuses on the qualities of the transgressing action itself as the locus of the emotion. Shame can sometimes be about very local failures as well, for example, when a person is ashamed of her fleeting bad thought. However, such shameful thoughts usually accompany a further judgment like the following: "How can *I* have such a thought?" If no such judgment would arise for the person, then, I think, we have good reason to doubt if her emotion is shame at all. That is, the target of shame seems to be typically (though perhaps not always) wrapped up in some of our general character traits or dispositions, even when the same object can also be the target of guilt.

Given that the self-action contrast is the most dominant basis for differentiating shame from guilt, the most natural rejoinder to the intertwinedness argument is to use this contrast against inclusivism. This is surely not the only possible rejoinder. Many philosophers have appealed to other ways of contrasting shame with guilt in order to classify only guilt as an accountability-tied

attitude. For example, Wallace (1994) argues that guilt rather than shame is involved in holding oneself to an expectation or demand; and Darwall (2006) suggests that guilt rather than shame exhibits a "second-personal" standpoint. These views usually come with a fairly substantial position about why moral accountability is a significant concept for us. I will come back to some of these approaches later and address the significance of moral accountability in §6.4.

Now consider the following line of argument.

(1) The primary object of concern in guilt, but not in shame, is the particular action that constitutes wrongdoing.

(2) The only things we are ever directly morally accountable for are particular actions that constitute wrongdoings.

(3) An accountability-tied reactive attitude must have the same primary object of concern as what we are directly morally accountable for.

(4) Therefore, shame cannot be an accountability-tied reactive attitude, but guilt can.

Call this the *Action-Based Objection* to inclusivism. The thought is that, if moral accountability has particular actions as the primary focus, then we should think only the emotion that has particular actions as its primary focus—guilt but not shame—is an accountability-tied reactive attitude. I think the objection has some initial plausibility. In particular, the assumption that we are only ever directly accountable for actions is an intuitive and traditional view. After all, how can a person be morally accountable if not due to something that she *did*? How can a person be morally blameworthy if she has not *done anything wrong*? It appears to be a plausible view that we are only *indirectly accountable* for things other than actions, when those things are the reasonably foreseeable consequences of what we do.

But why restrict the scope of moral accountability in this way? One common answer is that only actions are *within our voluntary control*, in the sense that we can choose whether to perform some actions. This connects to another, related argument against inclusivism, directly from the assumption that voluntary control is a necessary condition for moral accountability. The argument goes as follows.

121

(1) Guilt is appropriate only if its target is within the person's voluntary control, but shame can be appropriate even if its target is beyond the person's voluntary control.

(2) The only things we are ever directly morally accountable for are those within our voluntary control.

(3) An accountability-tied reactive attitude must have its appropriate condition being restricted by the necessary conditions of direct moral accountability,

(4) Therefore, Shame cannot be an accountability-tied reactive attitude, but guilt can.

Call this the *Control-Based Objection* to inclusivism. Carlsson (2019) and Portmore (2019a, 2019b) have both discussed a similar line of argument to defend divisionism.[3] Again, I find the objection has strong initial appeal. (1) is backed up by both intuitive cases and empirical results. We frequently feel shame about our beliefs, desires, or characters, which do not seem to be what we can directly choose to alter. Nonetheless, it appears that shame about these targets can be appropriate. Tracy and Robins (2006) present evidence that "uncontrollable attributions for negative events" lead to greater shame than guilt, while "controllable attributions for negative events" lead to greater guilt than shame (1347). (2) is a traditional and widely accepted picture about accountability: we are morally accountable for things that we voluntarily choose. It has been suggested that only such things can make blame *fair* and can make us *deserve* being blamed and feeling painful emotions. (3) also appears to be a reasonable constraint on the accountability-tied reactive attitudes.

Nonetheless, I think both the *Action-Based Objection* and *Control-Based Objection* should ultimately be rejected. Though they make sense on the assumption that direct accountability is only about actions or that direct accountability is only about things within our voluntary control, both assumptions should be rejected. Recent work on moral responsibility has offered reasons for expanding the scope of direct moral accountability beyond particular actions and beyond the domain of voluntary control. A closer examination on these developments, I shall argue, in fact

---

[3]Carlsson appears to endorse the argument, but he does not specify if the relevant notion of control is voluntary control. Portmore explicitly claims that the argument only works if control means something weaker than voluntary control.

lends support to treating shame as an accountability-tied reactive attitude. Just as an important facet of our practices of holding people morally accountable is more action-directed and thereby more strongly associated with guilt, these practices have another important facet that is more self-directed and thereby more strongly associated with shame.

6.3   Expanding the Scope of Moral Accountability

There have been prominent defenses of the claim that direct moral accountability does not just involve actions or just involve things within our voluntary control. Theorists argue that the scope of direct moral accountability should be expanded to include, for example, *attitudes*, *character*, and *the self*.[4] I contend that, if these defenses are on the right track, then the *Action-Based Objection* and the *Control-Based Objection* are undermined, and there are instead good reasons to endorse inclusivism.

To start with, consider three cases as follows:

Sexist Attitude. Amanda is a physics professor. When she teaches Physics 101, she consciously holds the belief that, in her class, male students are much more talented than female students in physics. But she neither expresses this sexist belief nor performs any discriminatory action based on this sexist belief.

Cold Character. Barbara has a character of cold-heartedness. For example, she has strong dispositions to dismiss her friends' emotions as unimportant, to not care about her close family members, and to not sympathize with those people who suffer from hard circumstances. But she neither expresses this cold character nor performs any cold action based on this character.

Cruel Self. Claire is a cruel person, and, furthermore, she identifies with her cruelness. She desires to see innocent people suffer, and she wants this desire to be effective in her actions. But she neither expresses her cruel self nor performs any cruel action based on her cruel self.[5]

---

[4] See, e.g., Smith (2005, 2012), Sher (2006), Hieronymi (2008), Graham (2014), Adams (1985), Holroyd (2012), Taylor (1976), Westlund (2003). Though some of these authors do not use the term 'moral accountability', I think this is a fair representation of their view. For example, Sher, Hieronymi, and Graham talk in terms of blameworthiness, which is typically associated with accountability; though Smith (2005) talks primarily in terms of attributability, she extends the idea to accountability in her more recent work (2012).

[5] Here I assume a view like Frankfurt's (1971), according to which one's self involves one's second order volitions,

In all three cases, the agent has a substandard attitude, character, or self without further performing any substandard action. Many details of the cases are left open, so the agents surely can have some excuses that help them escape being blameworthy. But can there be some contexts where they are morally accountable for the attitude, character, or self, in the absence of morally accountable actions? That is, does the fact that they did not perform any actions on the basis of the attitude, character, or self *exclude* the possibility of accountability? My guess is that people's intuitions will be varied. Laypeople's pretheoretical intuitions will most likely accompany a sense of confusion: it may initially seem that Amanda, Barbara, and Claire can surely be blameworthy, but the fact that they have not performed any substandard actions may also generate hesitation for people's having any clear-cut pretheoretical intuitions about the cases.

Here is a possible reaction upon reflection: Amanda, Barbara, and Claire are blameworthy, but only *indirectly* so. That is, they are blameworthy for the attitude, character, or self only to the extent that they were blameworthy for certain actions they did before that have the attitude, character, or self as reasonably foreseeable downstream consequences. This is a reasonable implication to draw from the cases, and I will consider this line of thought later in §6.5.1.

My focus, however, will be on another kind of diagnosis, according to which Amanda, Barbara, and Claire can be *directly morally accountable*, and thereby directly blameworthy for the attitude, character, or self. Theorists have provided various reasons to support this kind of diagnosis. The first is an argument based on ordinary practices. Regardless of our intuitions about the above cases, as theorists observe, we do *in our practices* blame people for their attitudes even in the absence of blameworthy actions, including "attitudes like fear, contempt, admiration, guilt, envy, and resentment" (Smith 2005, 254), beliefs and judgments (Hieronymi 2008), self-centered attitudes (Adams 1985), disrespect (Graham 2014), and even implicit attitudes (Holroyd 2012); character and traits (Sher 2006); and a person's self (Taylor 1976; Westlund 2003). These blaming practices, just as blaming practices based on actions, can be in either emotional forms, like

that is, those desires one desires to have and to be motivated by. This is not essential to my argument, and I expect that similar cases and arguments can be developed even if one adopts alternative reasonable views of the self.

resentment and indignation, or behavioral forms, like the outward manifestations of resentment and indignation. Though our blaming practices are fallible evidence for blameworthiness, they are evidence nonetheless, and a picture that renders inappropriate this wide range of blaming practices is hard to defend. The second is an argument based on quality of will, which plays a particularly important role in the defenses of responsibility for attitudes by Smith (2005) and Hieronymi (2008). They contend that attitudes can indicate a flaw in the extent to which one has *proper regard* for another person's moral standings and interests. This, Smith and Hieronymi suggest, is a sufficient basis for holding resentment or indignation toward people like Amanda. A similar line of argument can be extended to cases involving character and self too: to the extent that they indicate, or even constitute, a substandard quality of will, they are the proper targets of moral accountability. A third, related argument concerns the connection between responsibility for action and responsibility for attitude, character, and self. Sher (2006) highlights this tight connection by pointing to scenarios where an action's badness is a "near-inevitable" result of her bad character—which, according to Sher, suggests that she should be blamed for the character as well (2006, 65). The argument can be extended to cases involving attitudes and self: there are scenarios where an action's badness is a near-inevitable result of one's substandard attitudes, or one's substandard self, and, as a result, it will appear arbitrary to *categorically* exclude attitudes and self from the domain of direct moral accountability.

To summarize: We have good reasons to expand the scope of direct moral accountability to go beyond actions and include attitudes, character, and self, because (i) it is hard to defend a wholesale rejection of the wide range of blaming practices about things other than actions; (ii) attitudes, character, and self can indicate substandard qualities of will that some argue are sufficient for falling in the domain of moral accountability; and (iii) attitudes, character, and self can be so tightly connected to actions that drawing a sharp boundary of direct accountability in between is hardly plausible.

I think that this *expansive view on moral accountability*, the view that we are sometimes

directly accountable for things other than actions, including attitudes, character, or self, is justified. But it is fair to say this remains controversial. Still, let us take the claim seriously for the moment, and consider the further question: *If* we are directly morally accountable for attitudes, character, and self, what does this entail about reactive attitudes? The first upshot is that the *Action-Based Objection* to inclusivism is no longer sound, since it is not the case that we are only ever directly morally accountable for actions. Instead, we can be directly morally accountable for attitudes, character, and self. All those are typically more self-focused than action-focused. This is obvious in the case of character and self, but also rings true in the case of attitudes. Shame about an attitude is typically (though perhaps not always) connected with the thought that the attitude is likely to indicate something negative about one's character, self, or commitments; for instance, shame about one's racist attitude is typically connected with a thought that one is likely to lack a strong enough commitment against racism. The second upshot is that the *Control-Based Objection* to inclusivism is also going to be undermined. This is because attitudes, character, and the self are rarely within our voluntary control. Indeed, some defenders of the expansive view (e.g., Hieronymi, Sher) explicitly reject that any control condition is necessary for moral blameworthiness. The thought is that the kind of indication of flaws in quality of will is sufficient for moral blameworthiness, regardless of control. But another possibility is that accountability still requires control, but in a sense weaker than voluntary control. This may be a kind of "rational control" (Smith 2005, 265) defined by whether something depends on one's rational judgments, or "reasons-responsiveness" control defined by whether one or one's mechanism has a disposition to recognize and act according to the relevant reasons (e.g., Fischer & Ravizza 1998). Either way, the requirement of voluntary choice in moral accountability ends up being too demanding. We can then agree with the *Control-Based Objection* that the appropriateness condition of shame does not require voluntary control but maintain that a weaker control condition—probably that of rational control or reasons-responsiveness—still restricts the conditions under which shame, or a subcategory of shame, is appropriate. As an example, one may argue shame about an event

126

is appropriate, in the sense that we have pro tanto reasons to feel shame, only when whether the event occurs depends on our rational judgments (in Smith's sense). So shame about a racist thought can be appropriate on the assumption that whether one has that thought depends on one's rational judgments, whereas we do not even have reasons to feel shame about some of our bodily features, which we lack rational control over.

Further, I think the expansive view on moral accountability in fact provides more direct support for inclusivism. Consider the following, widely agreed connection between accountability and self-directed reactive attitudes:

> If a person is negatively morally accountable for *X*, then it is appropriate for her to hold reactive attitudes toward herself with regard to *X*.

In cases where *X* is an action, it makes sense to take guilt to be the relevant reactive attitude. But if *X* is an attitude, character, or self, then I am inclined to think that shame may be better as a candidate emotion to fill in the conditional. Consider the three earlier cases again. It seems to me that the natural and paradigmatic emotion to feel for Amanda is shame about her sexist thought, for Barbara is shame about her cold character, and for Claire is shame about her cruel self. Guilt, by contrast, seems much less natural. How can Amanda feel guilt about a thought, Barbara feel guilt about a character, Claire feel guilt about her self, without their feeling guilt about any particular actions? This intuition vindicates the self-action contrast. Guilt is indeed primarily directed toward actions, and when it comes to attitudes, character, and self, the primary object of concern is more about how they reflect the person as *who she is*, which is better captured by shame rather than guilt.

The thought is that there is a useful division of labor between shame and guilt that we can draw. Both concern moral accountability, but guilt is more about the aspect in which one is accountable for particular actions, whereas shame is more about the aspect in which one's being accountable for things—attitudes, character, self, and perhaps some actions—reflects its root in some substandard features of the agent's self. Both facets are crucial for practices involving moral

accountability. Though it may seem that moral accountability is always indexed to a particular substandard action, an actual or potential substandard self quite often falls in the scope of moral accountability as well. In fact, one common way for an individual to downplay a charge of blame is exactly to acknowledge her fault in the particular action, but not to in any way acknowledge the fault in her attitudes, character, or self. The above picture is desirable because it draws a clear division of labor between guilt and shame such that the two emotions play distinct yet interacting roles in our responsibility practices.

To recap, the argument in the current section can be formulated as follows:

The Expansion Argument for Inclusivism

(1) The expansive view on moral accountability is correct.

(2) If the expansive view on moral accountability is correct, then we have good reasons to endorse inclusivism.

(3) Thus, we have good reasons to endorse inclusivism.

One may respond by either rejecting (1) or rejecting (2), and I will consider the main objections along those lines in §6.5. It is worth noting that the *Expansion Argument* goes hand in hand with the *Intertwinedness Argument*. The intertwinedness argument, as stated in the last section, is only convincing when we are shown that the familiar ways in which shame and guilt are differentiated fail to justify categorizing shame and guilt differently. This is what the expansion argument partly aims to show: the most familiar differentiations between shame and guilt, in terms of either the self-action contrast or voluntary control, would not justify different categorizations between shame and guilt, but rather show that shame and guilt are in charge of different aspects of our accountability practices. The intertwinedness argument is thus a convincing move with this supplementation, since, unlike in the case of resentment and fury, shame and guilt do not seem to be differentiated in ways that can justify different categorizations of the two emotions. From a different perspective, we can also view the intertwinedness argument as supplementing the expansion argument. One rejoinder people may have against the expansion

argument is that, even though the most familiar ways of distinguishing shame from guilt do not contradict inclusivism, we may still speculate that there are other differentiations waiting to be found. However, the fact that shame and guilt are so deeply intertwined with each other renders this empirical speculation rather unlikely. More generally, we can combine the *Expansion Argument* and *Intertwinedness Argument* into a dilemma. On one hand, if one thinks there is no useful way of differentiating shame from guilt, then one already has no good resources to endorse exclusivism or divisionism, since one should already be convinced by the *Intertwinedness Argument*. On the other hand, if one wants to appeal to the familiar ways of differentiating shame from guilt to reject inclusivism, then one should be convinced by the *Expansion Argument* that those differentiations fail to justify categorizing shame and guilt differently. Either way, there are good reasons to endorse inclusivism, regardless of one's substantial view on the nature of shame and guilt.

6.4   The Significance of Moral Accountability

As mentioned earlier, some may still resist inclusivism on the basis of some fairly substantial conceptions of moral accountability. They may suggest that the normative significance of moral accountability comes from its relation to, for example, *desert* or *demands and expectations*, and only guilt rather than shame captures this significance. However, I am inclined to think that inclusivism is fully consistent with these conceptions of accountability. In fact, the earlier arguments may bring these conceptions under a new light by locating the division of conceptual labor between guilt and shame in illustrative ways.

Let's start with desert. It is widely believed that moral accountability entails desert of sanction or reward. It may be suggested that, because shame is not concerned with desert, it should not be an accountability-tied attitude. But why isn't shame concerned with desert? One answer is to again appeal to voluntary control (see Carlsson 2019): we only deserve adverse

treatments for what we have voluntary control of, but shame about *X* can be appropriate even when we cannot voluntarily control *X*. But this is just the *Control-Based Objection* to inclusivism discussed earlier and faces the same difficulties. Importantly, the inclusivist can reasonably maintain that accountability entails desert but deny that desert requires voluntary control. That is, we can pair both accountability and desert with weaker senses of control, like rational control or reasons-responsiveness. The desert-based objection then loses its force. The common examples against using control to restrict appropriate shame are usually those involving the lack of voluntary control, such as an attitude or a character that we cannot just choose or decide to alter. But it remains plausible that shame, or a subcategory of shame, cannot be appropriate if its object is in no way connected to our rational judgments or reason sensitivity at all. On this basis, we can maintain that shame, or a subcategory of shame, is in fact concerned with desert. It is also worth emphasizing that the deserved treatment should not be identified with harsh punishment. Though it is hardly plausible that Amanda, Barbara, and Claire deserve harsh punishment for their attitudes, character, and self, it is plausible that they deserve some adverse consequences, like other people's resentment, indignation, and the painful feelings of shame—of course, only proportional to how bad their attitudes, character, and self are.

Now turn to demands and expectations. One may think that the significance of moral accountability comes from those interpersonal demands or expectations that we hold people account to. Wallace defends exclusivism on the basis of this thought. Wallace appears to agree that shame is self-focused; he writes that when feeling shame one typically "sees oneself as being all of a piece" and "thinks of oneself as being thoroughly degraded" (1994, 241). But Wallace then suggests that "if one violates a demand that one holds oneself to, it will be very hard to think of oneself as all of a piece" (1994, 241). Assuming that reactive attitudes are instances of holding oneself to a demand, Wallace concludes that shame does not count as a reactive attitude.

However, it is unclear why there cannot be proper demands and expectations about the self-centered aspects of a person. On one way of understanding cases like Amanda, Barbara, and

130

Claire, they exemplify that we do in our blaming practices hold other people to account for what we demand or expect of their attitudes, character, and self. And the most natural ways in which we view ourselves in relation to these demands or expectations are in the form of shame rather than guilt. To say the least, it is unclear why the function of shame cannot be to bring the agent herself to see the force of these self-focused reciprocal demands or expectations. Watson (2014) has made a similar objection to Wallace. As Watson points out, even if there were "pure shame cultures," people in those cultures would still "be prone to a vivid sense of mutual expectations" and "recognize obligations to one another" (2014, 29). It is then reasonable to infer that, even in "our" culture, shame or an important subcategory of it can involve the kind of demands and expectations that is central to accountability practices. One may then ask Watson why we in fact have two distinct sets of emotional attitudes—guilt and shame—if they are both concerned with our interpersonal demands and expectations. My earlier arguments in this chapter help fill in this part of the story, since they show that there is still a useful division of labor between guilt and shame. Guilt involves holding oneself to more action-focused demands or expectations, whereas shame involves holding oneself to more self-focused demands or expectations, including those manifested in attitudes, character, and self.

I have considered two kinds of substantial conceptions of accountability—in terms of desert and interpersonal demands—and examined the plausibility of inclusivism in relation to these conceptions. These do not exhaust all possible conceptions, but the two examples provide important insights: First, theorists endorsing these conceptions frequently assume that accountability is only about actions or only about things within our voluntary control. Once these assumptions are rejected, it becomes clear that these conceptions can be consistent with, or even lend support to, inclusivism. Second, my earlier arguments for inclusivism can help illustrate how shame and guilt are concerned with different aspects of our accountability practices under these substantial conceptions of accountability.

6.5 Objections and Replies

6.5.1 Challenging the Expansive View

The expansive view on direct moral accountability remains controversial. An overall evaluation of this view, in contrast to the more restrictive view on the scope of direct moral accountability, can only be done by examining their various theoretical advantages and disadvantages and is beyond the scope of the current chapter. In this section, I offer a response to those who may reject the expansive view. The purpose is to show that even the more restrictive view on moral accountability may nonetheless support inclusivism.

To start with, note that, though the thesis that we are directly accountable for attitudes, character, and the self is fairly controversial, most theorists would agree that we can at least sometimes be *indirectly morally accountable* for an attitude, character, or self when and because (i) we are directly accountable for an action or a series of actions, and (ii) the formation of the attitude, character, or self is a reasonably foreseeable consequence of that action or that series of actions (cf., Fischer & Ravizza 1998; also see Brink 2021). But then we could make a similar inference from the expansive view on the scope of indirect moral accountability to inclusivism, by examining those cases where we *lack epistemic access to* the earlier actions that we are directly accountable for. Consider a case as follows:

> *Lazy Professor.* Dan is a mathematics professor. When he teaches Math 101, he consciously holds the belief that, in his class, male students are much more talented than female students in mathematics. But he neither expresses this sexist belief nor performs any discriminatory action based on this sexist belief. Further, Dan cannot recall what he did or decided not to do in the past that may have caused him more inclined to hold sexist beliefs. Still, it is reasonable to believe that Dan did something, such as skipping ethical trainings, that made him more likely to hold sexist beliefs.

What's special about this case is that the agent cannot really point to the thing that he was directly accountable for at an earlier time, though—let's stipulate—we have enough reasons to believe that there is such a thing that made him indirectly accountable for its downstream

consequence. Now consider the earlier biconditional between moral accountability and reactive attitudes again:

> If a person is negatively morally accountable for *X*, then it is appropriate for her to hold reactive attitudes toward herself with regard to *X*.

What are the appropriate negative reactive attitudes for Dan to feel? Perhaps we can say it is appropriate for Dan to feel guilt about what he did earlier, but this does not seem to fully capture his correct emotional response. Imagine Dan apologizes to the students by saying "I feel really guilty about what I did earlier—which, I am sorry, I cannot remember what that was…" It seems that such responses are far from satisfying, in the sense that they do not meet what we usually expect from agents in those situations. There should be something more. I contend that the "something more" here can be captured in their emotional response toward the consequence of their directly accountable actions. So Dan should feel some negative reactive attitude about his sexist belief, even if he could not remember what he did earlier.[6] We can then argue that, similar to the reasoning in §6.3, shame rather than guilt is the more natural and paradigmatic emotion for Dan to feel about his sexist attitude, and this provides a good reason to endorse inclusivism.

More generally, the link between reactive attitudes and moral responsibility should not be restricted to direct responsibility. As an analogy, consider the view that we are only directly morally responsible for choices and decisions, whereas responsibility for outward behaviors is a more derived notion. Those who believe this view surely should not think that the only reactive attitudes are the appropriate emotions toward choices and decisions; they will include the appropriate emotions about outward behaviors too. The upshot is that, *if* we are indirectly accountable for attitudes, then we still have good reasons to include shame as an accountability-tied reactive attitude. And the antecedent here is widely accepted even among those who reject

---

[6]Does this mean there is an asymmetry between guilt and shame, since guilt is appropriate only when one is *directly* morally accountable for an action? I am inclined to answer no. Suppose that Jim voluntarily decides to drink and drive but then involuntarily kills a pedestrian. If appropriate guilt is only about directly accountable actions, then it follows that Jim's proper guilt should be only about drinking and driving. But it seems to me that he should also feel guilty about the action of hitting the pedestrian—an action that he is only indirectly accountable for. Therefore, I tend to think we should reject the assumption that appropriate guilt only concerns directly accountable actions.

the expansive view on *direct* moral accountability.[7]

6.5.2   Challenging the Inference from the Expansive View to Inclusivism

What if one takes for granted the expansive view on direct moral accountability, but nonetheless rejects inclusivism? That is, why cannot one maintain that, in contrast to what I have suggested, we can feel guilty toward attitudes, character, and self? Why not think attitudes, or the processes involving forming attitudes, are themselves a kind of "mental action"?

Smith (2011), for example, claims that guilty attitudes are in fact fairly common phenomenon. She writes that,

> Perhaps we have caught ourselves taking secret pleasure in a close friend's uncharacteristic failure, or feeling distrustful of a loved one's fidelity, or viewing a stranger through the lens of an odious stereotype. Perhaps we have received the confidences of another with contempt, or have felt resentment rather than gratitude toward someone who has done us a kindness. It is quite common, in such cases, for people to say that they feel "guilty" about these thoughts and attitudes, even when they are quite sure that they have not been, and will not be, expressed or acted upon in any way. The feeling of guilt, it seems, attaches to the mere having of these thoughts and attitudes, and is not inhibited by the knowledge that these mental states are and will remain wholly private. (Smith 2011, 235)

My reply is as follows. Even if we do sometimes feel guilty about these thoughts, this still does not justify treating guilt and shame differently when it comes to categorizing them as reactive or non-reactive attitudes. Admittedly, the observation does pose a serious challenge to my claim that shame is the more natural and paradigmatic emotion to feel when it comes to attitudes, character, and the self; however, it still does not accomplish what one needs in order to argue against inclusivism, namely, a way of differentiation between shame and guilt that justifies treating only the latter as an accountability-tied reactive attitude. If anything, the observation almost strengthens the thought underlying the *Intertwinedness Argument*. Surely, many also feel

---

[7]However, it is worth noting that, on the picture discussed in the current subsection, shame is admittedly a less central reactive attitude than guilt, and its status as a reactive attitude is derived. This is not a concession that inclusivists would need to make if the expansive view on direct moral accountability is correct.

shame about the above attitudes—taking secret pleasure, distrusting a loved one, stereotyping strangers, etc., and, when it comes to attitudes, guilt and shame appear even more intertwined than in the case of actions. It is even more common than in actions that guilt and shame overlap, are hard to differentiate, and frequently convert to each other. We thereby still have not been shown that the intertwinedness can be unraveled to support exclusivism or divisionism.[8]

6.5.3   Challenging from the Dark Side of Shame

A final line of objection to my arguments involves appealing to certain allegedly unique, negative features of shame. For example, one may worry that shame is not an "autonomous" moral response, in the sense that it reflects not one's own values but rather the values of other people in one's society. The idea is that shame, even its "moral shame" or "agential shame" subcategory, seems to be strongly connected with one's social position, such as the social power and oppression that one is subject to. This would be an argument against inclusivism if moral accountability does not involve such social factors. However, there has been promising developments of the view that moral accountability has a crucial social dimension (see, e.g., Strawson 1962; McKenna 2018; Vargas 2018), which shame may be especially well-positioned to capture.

A different but related concern is that the moral relevance of shame is old news. A lot have been written on the positive moral value of shame, despite its apparently negative features.[9] However, the focus there is usually on the moral relevance of shame in general, rather than about the particular connection between shame and moral responsibility. Even when some theorists (e.g., Manion 2002; Deonna, Rodogno, & Teroni 2012) do discuss this particular connection, they do not address some of the more recent developments in the moral responsibility literature,

---

[8]We should also be cautious about drawing implications from those feelings we call "guilty thoughts." I suspect that guilty thoughts in fact share more similarities with paradigmatic shame attitudes than with paradigmatic guilt attitudes. See, e.g., Goffin & Cova (2019), who conduct experiments to show that guilty pleasures are "not so much about guilt (what people do) but about shame and embarrassment (who they are)" (1151).

[9]See, for example, Velleman (2001), Calhoun (2004), Nussbaum (2004), Deonna, Rodogno, & Teroni (2012), Deonna & Teroni (2011), Manion (2002), Mason (2010), Maibom (2010), Thomason (2015, 2018).

including the distinction between attributability and accountability, the expansive view on moral accountability, and the relation between direct and indirect accountability. My view is that we can establish a firmer defense of inclusivism by examining these specific choice-points in a theory of moral responsibility. Another way in which my arguments differ from the more general literature on the positive moral value of shame is this. My primary goal is to defend a connection between shame and moral accountability. I am inclined to think that defending this connection is in addition to, or even independent of, defending the positive moral value of shame. After all, there is the possibility that our practices involving holding people morally accountable do unfortunately have certain negative features and require significant revisions. We should not at the beginning of our theorizing simply assume that whatever emotions central to our responsibility practices must be conducive to our being better moral agents, or that such practices are flawless, or that they cannot be modified or improved (see Vargas 2013, 76–7). This possibility can be easily overlooked when we do not disentangle the goal of defending inclusivism on one hand, and that of defending the general moral value of shame on the other.

6.6   Concluding Remarks

I have presented novel support for including shame as a reactive attitude in the sense that concerns moral accountability, appealing to both the intertwinedness between guilt and shame, and the expansive view on the scope of moral accountability. This leads to various open questions. Where exactly should the scope of reactive attitudes be drawn? How about other self-directed and other-directed emotions such as pride, regret, and disgust? Does inclusivism about shame score better or worse once we consider these more general issues? I think they all lead to promising future projects, and I expect that the two aspects that I have focused on in the context of shame—the intertwinedness and interactions between different emotions, and the scope of moral accountability—will remain crucial for answering these broader questions.

Chapter 6, in part, has been accepted for publication (as "Shame and the Scope of Moral Accountability") in The Philosophical Quarterly, 2021, Wang, Shawn Tinghao. The dissertation/thesis author was the primary investigator and author of this paper.

Chapter 7

Conclusions

7.1 Two Themes

This dissertation had two major themes. The first theme is *pluralist functionalism about blame*: the view that blame's multiple functions are both definitive of what blame is and have normative implications about whether we ought to blame. I argued that blame's functions include protest, communication, and signaling. I proposed a functional definition of what blame is: blame is just the kind of practice that has a cluster of these three functions, and an individual token attitude or activity counts as blame just in case that it has a robust disposition to serve enough of the functions in this cluster. I also discussed a variety of normative implications from blame's functions: normative implications about the *eligibility*, *value*, and *propriety* of blame.

1. I suggested a context-specific strategy when using pluralist functionalism to draw implications about who are *eligible* targets of blame. If F is the only contextually significant function in a distinctive kind of context C, then, I proposed, eligible targets of blame must also be fitting targets when it comes to F. I treated this strategy as a potential modification of the communication argument, but it is worth noting that it can similarly apply to blame's non-communicative functions.

2. I argued that blame's multiple functions, when jointly realized, have a special *value*: it contributes to the realization of various first-order normative aims in an efficient and non-alienating way. I argued that this distinct value is what makes blame's multiple functions organically relate to each other. I also argued that the lack of access to this value underpins many cases of Blame Conflict, causing a special problem.

3. It may be the case that an individual is an eligible target of blame, and blaming her would bring about certain values, but we nonetheless *ought not to* blame her. We ought to blame her for her action only when it is *appropriate* to blame her for the action. I argued that pluralist functionalism has important implications about the propriety problem as well. I suggested that, if (i) one function of blame is to *F*, and (ii) *F* contributes to a morally good broader system, and (iii) *S* adopts *F* as one of her ends, then there is a consideration for *S* that speaks in favor of blaming when blame would fulfill this function. If this is the only relevant consideration, then *S* ought to blame; her blaming would be appropriate.

The second major theme of this dissertation involves *restrictionism about reactive attitudes*: the view that we should *restrict* and *reshape* the role of reactive attitudes in theorizing moral responsibility.

1. Their roles need to be *restricted*, in contrast to those approaches that take reactive attitudes to be the most fundamental in our responsibility practices. I argued that one such approach—the response-dependence account of moral responsibility—faces a granularity challenge. I suggested that one possible response involves restricting the metaphysical role of reactive attitudes.

2. The roles of reactive attitudes need to be *reshaped*, in the sense that we need to rethink the proper scope of such attitudes. I argued that shame should count as a reactive attitude, in contrast to the traditional wisdom. The upshot here is that the roles of some specific reactive attitudes like guilt have been overblown and should be properly restricted. Restrictionism

about reactive attitudes thus has two layers: globally, we should restrict the role of reactive attitudes as a general class; and locally, we should restrict the role of particular emotions like guilt that are standardly taken to be the paradigm of reactive attitudes.

Restrictionism about reactive attitudes goes hand in hand with pluralist functionalism about blame. On one hand, if blame's multiple functions—protest, communication, and signaling—are the key to understanding its nature and ethics, then reactive attitudes are going to be important as well, because they are often well-suited for serving such functions. But it also follows that reactive attitudes would have a limited and derived role, because they are important for theorizing moral responsibility ultimately *because and when* they serve, or are disposed to serve, certain functions. On the other hand, if the roles of reactive attitudes need to be restricted, then we might naturally wonder what can serve as the guidance for this restriction. Pluralist functionalism thus becomes appealing given its potential to providing guiding principles about when reactive attitudes' roles ought or ought not to be restricted.

7.2   Implications and Future Directions of the Work

I believe that pluralist functionalism provides a novel theoretical framework that we can use to examine a variety of issues about blame, moral responsibility, and moral emotions. There are many future directions that this project could take, some of which I have hinted at in the concluding sections of the earlier chapters. I will end this dissertation by sketching three related projects that I plan to pursue in the near future: a project about the proleptic function of blame (§7.2.1), a project about the social dimension of moral responsibility (§7.2.2), and a project about the individuation of different moral psychological phenomena (§7.2.3).

7.2.1 The Proleptic Function of Blame

I have discussed blame's signaling, protest, and communicative functions. Another possible function of blame that has received much attention in the literature is the *proleptic* function (Williams 1995; Fricker 2016, 2018; Bagley 2017; McGeer 2019). The proleptic function is a primarily forward-looking function, but it is different from other functions, such as blame's communicative function, in important ways. Blame communicates to people who are *already morally responsible agents*. By contrast, the idea of proleptic blame is that blame can target individuals who (i) do not strictly speaking possess the prerequisite features for being morally responsible agents yet, but (ii) our treating them *as if they already have those features* will lead to, in one way or another, the realization of their moral potential to develop those features in the future. Roughly speaking, blame in effect functions to help the wrongdoer "fake it until they make it" in the process of becoming fully morally responsible agents.

The proleptic view allegedly changes the landscape of blame rather significantly. For example, McGeer (2019) takes proleptic blame to come hand in hand with a new metaphysics of responsible agency, where responsible agents' abilities to respond to moral reasons are not merely static dispositions, but rather dynamic skills that take work and continual social feedback to develop and sustain. Fricker contends that proleptic blame has a "transformative" role (2018, 176), is a "useful piece of causal social construction" (2018, 165), and makes possible the process of generating shared reasons in the moral community.

However, there is something puzzling in the very idea of proleptic blame. It seems to involve some degree of *deception* or *pretending* that makes it questionable whether it is really a morally good function that properly contributes to the functioning of our first-order moral system. In addition, the claim that we can blame people who only have the *potential ability* to be morally responsible agents seems to open the door for some radical views, like the view that we can appropriately blame very young children.

I plan to investigate whether the proleptic function is really a function of blame, and,

if so, how it may or may not fit with the pluralist functionalist framework developed in this dissertation. A tentative hypothesis is that blame does have a proleptic function and that it can play an important role in the *transferring of agential capacities across contexts*. Let me explain. In Chapter 2, following Vargas (2013), I suggested that morally responsible agency is a context-sensitive concept: one may have responsible agency in a kind of context *C1* but not in another kind of context *C2*. My hypothesis is that blame functions proleptically by extending an individual's responsibility-relevant agential capacities from one kind of situations to another kind of situations. More accurately, the view is that, when blame properly fulfills its proleptic function, it targets an individual who, strictly speaking, does not possess the responsibility-relevant agential capacities with regards to situations of the kind *K1* yet, but (i) this individual has the adequate responsibility-relevant agential capacities with regards to situations of a different kind *K2*; (ii) the agential capacities at issue are transferable from situations of kind *K2* to situations of kind *K1*; and (iii) blaming the individual is an essential part of bringing her to recognize and initiate this process of transferring her agential capacities.

Though I find this cross-situational proleptic function of blame plausible, and I thereby think that an account of proleptic blame is promising, more work needs to be done to understand the relation between the proleptic function and the signaling, protest, and communicative functions. Do they form a function-cluster or not? Is the proleptic function reducible to any of these other functions? I plan to explore if adding the proleptic function to blame's function-cluster would require the modification of some claims in this dissertation. For example, I have argued that blame often has a value of playing a *non-alienating* role in shaping agency. But the proleptic function may seem fairly alienating; it involves some degree of deception so that the wrongdoer may feel she is distanced from participating in enough of the process of realizing blame's functions. This may then require us to modify the initial claim that blame is often non-alienating.

7.2.2 The Social Dimension of Moral Responsibility

Another implication of this dissertation concerns the social dimension of moral responsibility. The functions that I have focused on are *social* in a fairly thin sense: they concern the interaction between different persons in a society, in the form of protest, communication, and signaling. But one may wonder if blame has *social functions in a thicker sense*, such that the description of a function of blame may be incomplete without referring to some fairly specific social factors—for example, factors about oppression, marginalization, and privilege. I have briefly discussed a similar issue in §4.5.1: Does blame have functions such as marginalizing minorities, securing the established social hierarchies, etc.? The worry is closely related to the worry about morally perverse functions. Some candidates for blame's thick social functions are notoriously bad or oppressive: for example, one may argue that blame has a function of patronizing (see Connelly 2022), or a function of leading to unfair criticisms of the victim (sometimes called "victim blaming").[1] In §4.5.1, I have suggested some general responses to the objection from morally perverse functions, but I take it to be important work to understand the more specific forms of social privilege and oppression and how they may produce problematic functions of blame.

I plan to more systematically investigate these issues in the near future. In addition to clarifying the contents of blame's function-cluster, understanding blame's thick social functions also has important implications about the nature of morally responsible agency. Some theorists have recently suggested that morally responsibility has a crucial social dimension that has been overlooked in the traditional accounts. These theorists (e.g., McGeer 2012, 2019; Vargas 2013; McKenna 2018) develop instead *socially constituted* or *socially scaffolded* accounts of morally responsible agency. Their thesis is not merely that social factors have a *causal* impact on the development and sustaining of responsibility-relevant capacities; rather, they are committed to

---

[1]Another worry is that our judgments about morally responsible are often biased by social factors (see, e.g., Argetsinger 2022).

the stronger view that responsibility-relevant capacities are at least partly *constituted* by social factors, including but not limited to factors such as social deprivation, biases, oppression, and marginalization. However, one major problem with these approaches is that it is notoriously difficult to spell out what it exactly means to say that responsible agency is socially constituted. Pluralist functionalism has the potential of helping make sense of this claim. Remember that in Chapter 2, I suggested that whether one has responsible agency depends on whether blame would fit its contextually significant functions. If some of these functions are social in the fairly thick sense, then it is plausible that these functions themselves are socially constituted. Assuming that contextually significant functions are conceptually tied to responsible agency, it would follow that morally responsible agency is also socially constituted. This provides another motivation for further exploring the possible thick social functions of blame.

This project would also provide resources for understanding the role of reactive attitudes in the inquiry about the sociality of responsibility. More specifically, if blame turns out to have thick social functions, then we need to shift our study of reactive attitudes to the socially thick moral emotions. This is similar to the line of argument I presented in Chapter 6, where I contended that shame—an emotion that is strongly connected to our social practices—is a moral emotion relevant to moral responsibility. I plan to investigate if there are other socially thick emotions that capture important facets of our responsibility practices. This may require us to explore some emotions that may initially seem too "social" or not "autonomous" enough to count as reactive attitudes, such as contempt (Mason 2018) and even spite (Thomason 2020). This may also require us to look deeper into sociology and social theories, which may provide resources for understanding psychology, and may offer new moral emotions, when it comes to oppression, marginalization, and privilege.

144

7.2.3 Individuation of Moral Psychological Phenomena

I started this dissertation by asking how blame is different from various nearby practices. Let's come back to this question. According to pluralist functionalism, blame is defined by a cluster of functions including protest, communication, and signaling. So part of the answer to the initial question is that none of blame's nearby practices has the same functional profile; even when, for example, feeling sad has a signaling function and punishment has a communicative function, none of the nearby practices has all three functions that blame has. Another part of the answer is again the distinct value of blame's functions: they contribute to first-order normative aims in an efficient and non-alienating way.

For some of blame's nearby practices, it is fairly easy to see why they do not involve this kind of contribution. The "colder" responses to wrongdoings—punishment, threat, and warning—are usually so cold that they alienate people. People being punished and warned are less likely to feel personally involved and invested in the process of realizing the relevant moral aims; we often risk damaging our relationships when we issue these cold responses to those whom we love and care about; and when it comes to socially salient issues, these cold responses tend to alienate our potential allies. By contrast, the "milder" responses to wrongdoings—disappointment, frustration, and sadness— are usually too mild to make any difference. They may be efficient in realizing various *intrapersonal* moral aims, such as cultivating good dispositions and reaffirming one's values; but they are usually inefficient when it comes to the *interpersonal* moral aims. At least, they are usually not as efficient as blame; for example, blaming a friend for a mistake is, as a general matter, more likely than merely expressing one's frustration to promote changes in the friend's future behaviors.

None of these is supposed to mean that blame is the *best* response to norm violations. What counts as the best depends on the context. I discussed blame's contextually significant functions in Chapter 2. I am inclined to think that we can similarly talk about the contextually significant moral values or contextually significant first-order aims—those values and first-order

aims that should matter given the morally salient features of the circumstance that deserve our moral attention. Whether blame is the best response in a certain context, then, would depend on whether it is disposed to realize these first-order aims and values.

And even when it comes to the value of contributing to first-order aims in an efficient and non-alienating way, there are further issues to address. One issue concerns exactly *how efficient* blame is in realizing those aims. Evidence from empirical psychology and the social sciences will turn out to be relevant. Another issue concerns whether some alternative practice is also efficient and non-alienating. One potential candidate is the practice of *moral education*. Isn't moral education, at least when it is done in an ideal way, meant to be promoting certain values both with a good efficiency and with little to no alienation effect? Relatedly, how about the practice of *encouraging* appropriate attitudes?[2] It may turn out that these practices, like blame, also contribute to first-order normative aims in an efficient and non-alienating way. There are still two further possibilities of telling these practices apart, under the pluralist functionalism framework. First, it may be the case that they realize *different first-order normative aims* from blame. Second, it may be the case that they contribute to the same first-order normative aims in an efficient and non-alienating way, but through different mechanisms. We would then have to appeal to blame's particular function-cluster—which consists of protest, communication, and signaling—in order to distinguish blame from these alternative practices. I take it to be important future work to apply the pluralist functionalist framework to moral psychological phenomena in general, and to investigate how blame is similar to and different from other practices like moral education. It is also worth noting that moral education may itself involve some form of blame, perhaps the kind of proleptic blame discussed in §7.2.1—the kind of blame that aims to shape and improve children's (and adults') responsible agency.

---

[2]For reference, see Brink and Nelkin's (2022) discussion on moral education and functionalism about blame.

References

Adams, R. M. (1985). Involuntary sins. *The Philosophical Review*, *94*(1), 3–31.

Argetsinger, H. (2022). *Killing in broad daylight: Social status and epistemic distortion.* (Unpublished Manuscript)

Arneson, R. (2003). The Smart theory of moral responsibility and desert. In S. Olsaretti (Ed.), *Desert and justice* (pp. 233–258). Oxford University Press.

Arriaga, X. B., & Rusbult, C. E. (1998). Standing in my partner's shoes: Partner perspective taking and reactions to accommodative dilemmas. *Personality and Social Psychology Bulletin*, *24*(9), 927–948.

Bagley, B. (2017). Properly proleptic blame. *Ethics*, *127*(4), 852–882.

Barrett, L. F. (2006). Valence is a basic building block of emotional life. *Journal of Research in Personality*, *40*(1), 35–55.

Beglin, D. (2018). Responsibility, libertarians, and the "facts as we know them": A concern-based construal of Strawson's reversal. *Ethics*, *128*(3), 612–625.

Beglin, D. (2020). Two Strawsonian strategies for accounting for morally responsible agency. *Philosophical Studies*, *177*, 2341–2364.

Bigelow, J., & Pargetter, R. (1987). Functions. *Journal of Philosophy*, *84*, 181–196.

Boorse, C. (1976). Wright on functions. *Philosophical Review*, *85*, 70–86.

Bouchard, F. (2013). How ecosystem evolution strengthens the case for function pluralism. In P. Huneman (Ed.), *Function: Selection and mechanisms* (pp. 83–95). Springer.

Boxill, B. (1976). Self-respect and protest. *Philosophy & Public Affairs*, *6*, 58–69.

Boyd, R. N. (1988). How to be a moral realist. In G. Sayre-McCord (Ed.), *Essays on moral realism* (pp. 181–228). Cornell University Press.

Boyd, R. N. (1999a). Homeostasis, species, and higher taxa. In R. A. Wilson (Ed.), *Species: New*

interdisciplinary essays (pp. 141–185). MIT Press.

Boyd, R. N. (1999b). Kinds, complexity and multiple realization. *Philosophical Studies*, *95*, 67–98.

Boyd, R. N. (2003a). Finite beings, finite goods: The semantics, metaphysics and ethics of naturalist consequentialism, part I. *Philosophy and Phenomenological Research*, *LXVI*, 505–553.

Boyd, R. N. (2003b). Finite beings, finite goods: The semantics, metaphysics and ethics of naturalist consequentialism, part II. *Philosophy and Phenomenological Research*, *LXVII*, 24–47.

Brink, D. O. (1994). Moral conflict and its structure. *The Philosophical Review*, *103*(2), 215–247.

Brink, D. O. (2018). Partial responsibility and excuse. In H. Hurd (Ed.), *Moral puzzles and legal perplexities* (pp. 39–59). Cambridge University Press.

Brink, D. O. (2021). *Fair opportunity and responsibility*. Oxford University Press.

Brink, D. O., & Nelkin, D. K. (2013). Fairness and the architecture of responsibility. In D. Shoemaker (Ed.), *Oxford studies in agency and responsibility* (Vol. 1, pp. 284–313). Oxford University Press.

Brink, D. O., & Nelkin, D. K. (2022). The nature and significance of blame. In M. Vargas & J. Doris (Eds.), *Oxford handbook of moral psychology* (pp. 177–196). Oxford University Press.

Buller, D. J. (1998). Etiological theories of function: A geographical survey. *Biology and Philosophy*, *13*, 505–527.

Bykvist, K. (2009). No good fit: Why the fitting attitude analysis of value fails. *Mind*, *118*(469), 1–30.

Calhoun, C. (1989). Responsibility and reproach. *Ethics*, *99*(2), 389–406.

Calhoun, C. (1999). Moral failure. In C. Card (Ed.), *On feminist ethics and politics* (pp. 81–99). University Press of Kansas.

Calhoun, C. (2004). An apology for moral shame. *Journal of Political Philosophy*, *12*(2), 127–146.

Calhoun, C. (2016). *Moral aims: Essays on the importance of getting it right and practicing morality with others*. Oxford University Press.

Campbell, J. (2017). P. F. Strawson's free will naturalism. *International Journal for the Study of*

Skepticism, *7*, 26–52.

Carlsson, A. B. (2017). Blameworthiness as deserved guilt. *Journal of Ethics*, *21*(1), 89–115.

Carlsson, A. B. (2019). Shame and attributability. In D. Shoemaker (Ed.), *Oxford studies in agency and responsibility* (Vol. 6, pp. 112–139). Oxford University Press.

Clark, C. J. (2022). The blame efficiency hypothesis: An evolutionary framework to resolve rationalist and intuitionist theories of moral condemnation. In T. Nadelhoffer & A. Monroe (Eds.), *Advances in experimental philosophy of free will and responsibility* (pp. 27–44). Bloomsbury Publishing.

Clarke, R. (2013). Some theses on desert. *Philosophical Explorations*, *16*(2), 153–164.

Clarke, R. (2016). Moral responsibility, guilt, and retributivism. *The Journal of Ethics*, *20*(1), 121–137.

Coates, D. J., & Swenson, P. (2013). Reasons-responsiveness and degrees of responsibility. *Philosophical Studies*, *165*(2), 629–645.

Coates, D. J., & Tognazzini, N. A. (2013). The contours of blame. In D. J. Coates & N. A. Tognazzini (Eds.), *Blame: Its nature and norms* (pp. 3–26). Oxford University Press.

Cogley, Z. (2013). The three-fold significance of the blaming emotions. In D. Shoemaker (Ed.), *Oxford studies in agency and responsibility* (Vol. 1, pp. 205–224). Oxford University Press.

Connelly, K. (2022). *Blame and patronizing.* (Unpublished Manuscript)

Craver, C. (2001). Role functions, mechanisms, and hierarchy. *Philosophy of Science*, *68*, 53–74.

Cummins, R. (1975). Functional analysis. *The Journal of Philosophy*, *72*(20), 741–765.

Dancy, J. (2004). Enticing reasons. In R. J. Wallace, P. Pettit, S. Scheffler, & M. Smith (Eds.), *Reasons and value: Themes from the moral philosophy of Joseph Raz* (pp. 99–118). Oxford University Press.

Darwall, S. (2006). *The second-person standpoint: Morality, respect, and accountability*. Harvard University Press.

Darwall, S. (2015). *Second-personal attitudes of the heart.* Retrieved from PEA Soup: https://peasoup.typepad.com/peasoup/2015/04/second-personal-attitudes-of-the-heart-by-featured-philosopher-steve-darwall.html.

Darwall, S. (2016). Love's second-personal character: Holding, beholding, and upholding. In E. E. Kroeker & K. Schaubroeck (Eds.), *Love, reason, and morality* (pp. 93–109). Routledge.

Darwall, S. (2017). Trust as a second-personal attitude (of the heart). In P. Faulkner & T. Simpson (Eds.), *The philosophy of trust* (pp. 35–50). Oxford University Press.

Darwall, S. (2019). Gratitude as a second-personal attitude (of the heart). In R. Roberts & D. Telech (Eds.), *The moral psychology of gratitude* (pp. 139–159). Rowman & Littlefield.

Deonna, J., & Teroni, F. (2011). Is shame a social emotion? In A. K. Ziv, K. Lehrer, & H. B. Schmid (Eds.), *Self-evaluation: Affective and social grounds of intentionality* (pp. 193–212). Springer.

Deonna, J. A., Rodogno, R., & Teroni, F. (2012). *In defense of shame: The faces of an emotion*. Oxford University Press.

Duff, A. (2001). *Punishment, communication, and community*. Cornell University Press.

Fassio, D. (2011). Belief, correctness and normativity. *Logique et Analyse*, *54*(216), 471–486.

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

Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge University Press.

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

Fodor, J. A. (1968). *Psychological explanation*. Random House.

Frankfurt, H. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy*, *68*, 5–20.

Fricker, M. (2016). What's the point of blame? A paradigm based explanation. *Noûs*, *50*(1), 165–183.

Fricker, M. (2018). Ambivalence about forgiveness. *Royal Institute of Philosophy Supplement*, *84*, 161–185.

Funk, F., McGeer, V., & Gollwitzer, M. (2014). Get the message: Punishment is satisfying if the transgressor responds to its communicative intent. *Personality and Social Psychology Bulletin*, *40*(8), 986–997.

Garson, J. (2016). *A critical overview of biological functions*. Springer International Publishing.

Garson, J. (2019). *What biological functions are and why they matter*. Cambridge University Press.

Goffin, K., & Cova, F. (2019). An empirical investigation of guilty pleasures. *Philosophical*

Psychology, *32*(7), 1129–1155.

Gowans, C. (1996). Moral theory, moral dilemmas, and moral responsibility. In H. E. Mason (Ed.), *Moral dilemmas and moral theory* (pp. 199–215). Oxford University Press.

Graham, P. A. (2014). A sketch of a theory of moral blameworthiness. *Philosophy and Phenomenological Research*, *88*(2), 388–409.

Hardcastle, V. G. (2002). On the normativity of functions. In A. Ariew, R. Cummins, & M. Perlman (Eds.), *Functions: New essays in the philosophy of psychology and biology* (pp. 144–156). Oxford University Press.

Helm, B. (2014). Trust as a reactive attitude. In D. Shoemaker & N. Tognazzini (Eds.), *Oxford studies in agency and responsibility* (Vol. 2, pp. 187–214). Oxford University Press.

Herman, B. (1993). *The practice of moral judgment*. Harvard University Press.

Hieronymi, P. (2001). Articulating an uncompromising forgiveness. *Philosophy and Phenomenological Research*, *62*(3), 529–555.

Hieronymi, P. (2004). The force and fairness of blame. *Philosophical Perspectives*, *18*(1), 115–148.

Hieronymi, P. (2008). Responsibility for believing. *Synthese*, *161*(3), 357–373.

Hieronymi, P. (2019). I'll bet you think this blame is about you. In D. Shoemaker (Ed.), *Oxford studies in agency and responsibility* (Vol. 5, pp. 60–87). Oxford University Press.

Holroyd, J. (2012). Responsibility for implicit bias. *Journal of Social Philosophy*, *43*(3), 274–306.

Kaiserman, A. (2018). More of a cause: Recent work on degrees of causation and responsibility. *Philosophy Compass*, *13*, e12498.

Keltner, D., & Buswell, B. N. (1996). Evidence for the distinctness of embarrassment, shame, and guilt: A study of recalled antecedents and facial expressions of emotion. *Cognition and Emotion*, *10*(2), 155–171.

Kitcher, P. (1993). Function and design. *Midwest Studies in Philosophy*, *18*, 379–397.

Knobe, J., & Doris, J. M. (2010). Responsibility. In J. Doris & the Moral Psychology Research Group (Eds.), *The moral psychology handbook* (pp. 321–354). Oxford University Press.

Kogelmann, B., & Wallace, R. H. (2018). Moral diversity and moral responsibility. *Journal of the American Philosophical Association*, *4*(3), 371–389.

Kumar, V. (2015). Moral judgment as a natural kind. *Philosophical Studies*, *172*(11), 2887–2910.

Larsen, J. T., McGraw, A. P., & Cacioppo, J. T. (2001). Can people feel happy and sad at the same time? *Journal of Personality and Social Psychology*, *81*(4), 684–696.

Lewis, D. (1989). Dispositional theories of value. *Proceedings of the Aristotelian Society, Supplementary Volumes*, *63*, 113–137.

Lewis, H. B. (1971). *Shame and guilt in neurosis*. International Universities Press.

Lewis, M. (2008). Self-conscious emotions: Embarrassment, pride, shame, and guilt. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 742–756). Guilford Press.

Macnamara, C. (2013). Taking demands out of blame. In D. J. Coates & N. Tognazzini (Eds.), *Blame: Its nature and norms* (pp. 141–161). Oxford University Press.

Macnamara, C. (2015a). Blame, communication, and morally responsible agency. In R. Clarke, M. McKenna, & A. M. Smith (Eds.), *The nature of moral responsibility: New essays* (pp. 211–236). Oxford University Press.

Macnamara, C. (2015b). Reactive attitudes as communicative entities. *Philosophy and Phenomenological Research*, *90*(3), 546–569.

Mahtani, A. (2020). The dispositional account of credence. *Philosophical Studies*, *177*, 727–745.

Maibom, H. L. (2010). The descent of shame. *Philosophy and Phenomenological Research*, *80*(3), 566–594.

Manion, J. C. (2002). The moral relevance of shame. *American Philosophical Quarterly*, *39*(1), 73–90.

Manley, D., & Wasserman, R. (2008). On linking dispositions and conditionals. *Mind*, *117*(465), 59–84.

Mason, M. (2010). On shamelessness. *Philosophical Papers*, *39*(3), 401–425.

Mason, M. (2017). Contempt: At the limits of reactivity. In M. Mason (Ed.), *The moral psychology of contempt* (pp. 173–192). Rowman and Littlefield.

McGeer, V. (2012). Co-reactive attitudes and the making of moral community. In R. Langdon & C. Mackenzie (Eds.), *Emotions, imagination, and moral reasoning* (pp. 299–326). Psychology Press.

McGeer, V. (2013). Civilizing blame. In D. J. Coates & N. A. Tognazzini (Eds.), *Blame: Its nature and norms* (pp. 162–188). Oxford University Press.

McGeer, V. (2019). Scaffolding agency: A proleptic account of the reactive attitudes. *European Journal of Philosophy*, *27*(2), 301–323.

McGeer, V., & Funk, F. (2017). Are "optimistic" theories of criminal justice psychologically feasible? The probative case of civic republicanism. *Criminal Law and Philosophy*, *11*(3), 523–544.

McHugh, C. (2012). The truth norm of belief. *Pacific Philosophical Quarterly*, *93*(1), 8–30.

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

McKenna, M. S. (2014). Defending conversation and responsibility: Reply to Dana Nelkin and Holly Smith. *Philosophical Studies*, *171*(1), 73–84.

McKenna, M. S. (2016a). Quality of will, private blame and conversation: Reply to Driver, Shoemaker, and Vargas. *Criminal Law and Philosophy*, *10*(2), 243–263.

McKenna, M. S. (2016b). Reasons-responsiveness theories of freedom. In K. Timpe, M. Griffith, & N. Levy (Eds.), *The Routledge companion to free will* (pp. 49–62). Routledge.

McKenna, M. S. (2018). Power, social inequities, and the conversational theory of moral responsibility. In K. Hutchison, C. Mackenzie, & M. Oshana (Eds.), *Social dimensions of moral responsibility* (pp. 38–58). Oxford University Press.

McLaughlin, P. (2001). *What functions explain: Functional explanation and self-reproducing systems*. Cambridge University Press.

Mele, A. R. (1992). *Springs of action*. Oxford University Press.

Mele, A. R. (2009). *Effective intentions: The power of conscious will*. Oxford University Press.

Menges, L. (2017a). The emotion account of blame. *Philosophical Studies*, *174*(1), 257–273.

Menges, L. (2017b). Grounding responsibility in appropriate blame. *American Philosophical Quarterly*, *54*(1), 15–24.

Millikan, R. G. (1989). In defense of proper functions. *Philosophy of Science*, *56*, 288–302.

Nagel, E. (1961). *The structure of science*. Hackett.

Nanay, B. (2010). A modal theory of function. *Journal of Philosophy*, *107*, 412–431.

Neander, K. (1991a). Functions as selected effects: The conceptual analyst's defense. *Philosophy of Science*, *58*, 168–184.

Neander, K. (1991b). The teleological notion of "function". *Australasian Journal of Philosophy*,

69(4), 454–468.

Nelkin, D. K. (2011). *Making sense of freedom and moral responsibility*. Oxford University Press.

Nelkin, D. K. (2016a). Accountability and desert. *The Journal of Ethics*, *20*, 173–189.

Nelkin, D. K. (2016b). Difficulty and degrees of moral praiseworthiness and blameworthiness. *Noûs*, *50*(2), 356–378.

Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme*, *17*(2), 141–161.

Nichols, S. (2015). *Bound: Essays on free will and responsibility*. Oxford University Press.

Niedenthal, P. M., Tangney, J. P., & Gavanski, I. (1994). "If only I weren't" versus "if only I hadn't": Distinguishing shame and guilt in counterfactual thinking. *Journal of Personality and Social Psychology*, *67*(4), 585–595.

Nowell-Smith, P. (1948). Freewill and moral responsibility. *Mind*, *57*(225), 45–61.

Nussbaum, M. C. (2004). *Hiding from humanity: Disgust, shame, and the law*. Princeton University Press.

Nussbaum, M. C. (2016). *Anger and forgiveness: Resentment, generosity, justice*. Oxford University Press.

Owens, D. J. (2003). Does belief have an aim? *Philosophical Studies*, *115*(3), 283–305.

Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition*, *107*(1), 179–217.

Pereboom, D. (2014). *Free will, agency, and meaning in life*. Oxford University Press.

Pickard, H. (2013). Irrational blame. *Analysis*, *73*(4), 613–626.

Polger, T. W. (2021). *Functionalism*. Retrieved from IEP: https://www.iep.utm.edu/.

Portmore, D. W. (2019a). Control, attitudes, and accountability. In D. Shoemaker (Ed.), *Oxford studies in agency and responsibility* (Vol. 6, pp. 7–32). Oxford University Press.

Portmore, D. W. (2019b). Desert, control, and moral responsibility. *Acta Anal*, *34*, 407–426.

Portmore, D. W. (2022). A comprehensive account of blame: Self-blame, non-moral blame, and blame for the non-voluntary. In A. B. Carlsson (Ed.), *Self-blame and moral responsibility* (pp. 48–76). Cambridge University Press.

Queloz, M. (2021). The self-effacing functionality of blame. *Philosophical Studies*, *178*, 1361–1379.

Railton, P. (1986). Facts and values. *Philosophical Topics*, *14*(2), 5–31.

Rawls, J. (1971). *A theory of justice*. Harvard University Press.

Raz, J. (1999). *Engaging reason: On the theory of value and action*. Oxford University Press.

Reis-Dennis, S. (2019). Anger: Scary good. *Australasian Journal of Philosophy*, *97*(3), 451–464.

Reis-Dennis, S. (2021). Rank offence: The ecological theory of resentment. *Mind*, *130*(520), 1233–1251.

Rosen, G. (2015). The alethic conception of moral responsibility. In R. Clarke, M. McKenna, & A. Smith (Eds.), *The nature of moral responsibility: New essays* (pp. 65–89). Oxford University Press.

Ross, W. D. (1930). *The right and the good*. Oxford University Press.

Russell, P. (1992). Strawson's way of naturalizing responsibility. *Ethics*, *102*, 287–302.

Russell, P. (2004). Responsibility and the condition of moral sense. *Philosophical Topics*, *32*(1), 287–305.

Scanlon, T. M. (2008). *Moral dimensions: Permissibility, meaning, and blame*. Harvard University Press.

Scanlon, T. M. (2015). Forms and conditions of responsibility. In R. Clarke, M. McKenna, & A. Smith (Eds.), *The nature of moral responsibility: New essays* (pp. 89–111). Oxford University Press.

Schroeder, M. (2010). Value and the right kind of reason. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics* (Vol. 5, pp. 25–55). Oxford University Press.

Shah, N. (2003). How truth governs belief. *Philosophical Review*, *112*(4), 447–482.

Shah, N., & Velleman, J. D. (2005). Doxastic deliberation. *Philosophical Review*, *114*(4), 497–534.

Sher, G. (2006). *In praise of blame*. Oxford University Press.

Shoemaker, D. (2007). Moral address, moral responsibility, and the boundaries of the moral community. *Ethics*, *118*(1), 70–108.

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

Shoemaker, D. (2017). Response-dependent responsibility; or, a funny thing happened on the way to blame. *Philosophical Review*, *126*(4), 481–527.

Shoemaker, D. (2019). Hurt feelings. *The Journal of Philosophy*, *116*, 125–148.

Shoemaker, D., & Vargas, M. (2021). Moral torch fishing: A signaling theory of blame. *Noûs*, *55*(3), 581–602.

Smart, J. J. C. (1961). Free-will, praise, and blame. *Mind*, *70*, 291–306.

Smith, A. M. (2005). Responsibility for attitudes: Activity and passivity in mental life. *Ethics*, *115*, 236–271.

Smith, A. M. (2007). On being responsible and holding responsible. *The Journal of Ethics*, *11*(4), 465–484.

Smith, A. M. (2008). Control, responsibility, and moral assessment. *Philosophical Studies*, *138*, 367–392.

Smith, A. M. (2011). Guilty thoughts. In C. Bagnoli (Ed.), *Morality and the emotions* (pp. 235–256). Oxford University Press.

Smith, A. M. (2012). Attributability, answerability, and accountability: In defense of a unified account. *Ethics*, *122*(3), 575–589.

Smith, A. M. (2013). Moral blame and moral protest. In D. J. Coates & N. A. Tognazzini (Eds.), *Blame: Its nature and norms* (pp. 27–48). Oxford University Press.

Smith, M. (1994). *The moral problem*. Wiley-Blackwell.

Srinivasan, A. (2018). The aptness of anger. *Journal of Political Philosophy*, *26*(2), 123–144.

Steglich-Petersen, A. (2006). No norm needed: On the aim of belief. *Philosophical Quarterly*, *56*(225), 499–516.

Strawson, P. (1962). Freedom and resentment. *Proceedings of the British Academy*, *48*, 1–25.

Talbert, M. (2008). Blame and responsiveness to moral reasons: Are psychopaths blameworthy? *Pacific Philosophical Quarterly*, *89*, 516–535.

Talbert, M. (2012). Moral competence, moral blame, and protest. *Journal of Ethics*, *16*(1), 89–109.

Tangney, J. P. (1992). Situational determinants of shame and guilt in young adulthood. *Personality and Social Psychology Bulletin*, *18*, 199–206.

Tangney, J. P. (1995). Recent empirical advances in the study of shame and guilt. *American Behavioral Scientist*, *38*, 1132–1145.

Tangney, J. P., & Dearing, R. L. (2002). *Shame and guilt*. The Guilford Press.

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual Review of Psychology*, *58*(1), 345–372.

Taylor, C. (1976). Responsibility for self. In A. O. Rorty (Ed.), *The identities of persons* (pp. 281–299). University of California Press.

Telech, D. (2020). Demanding more of Strawsonian accountability theory. *European Journal of Philosophy*, *28*(4), 926–941.

Tessman, L. (2014). *Moral failure: On the impossible demands of morality*. Oxford University Press.

Thomason, K. K. (2015). Shame, violence, and morality. *Philosophy and Phenomenological Research*, *91*(1), 1–24.

Thomason, K. K. (2018). *Naked: The dark side of shame and moral life*. Oxford University Press.

Thomason, K. K. (2020). I'll show you: Spite as a reactive attitude. *The Monist*, *103*(2), 163–175.

Todd, P. (2016). Strawson, moral responsibility, and the "order of explanation": An intervention. *Ethics*, *127*(1), 208–240.

Tognazzini, N. A. (2013). Blameworthiness and the affective account of blame. *Philosophia*, *41*(4), 1299–1312.

Tracy, J. L., & Robins, R. W. (2006). Appraisal antecedents of shame and guilt: Support for a theoretical model. *Personality and Social Psychology Bulletin*, *32*(10), 1339–1351.

Vargas, M. (2004). Responsibility and the aims of theory: Strawson and revisionism. *Pacific Philosophical Quarterly*, *85*(2), 218–241.

Vargas, M. (2013). *Building better beings: A theory of moral responsibility*. Oxford University Press.

Vargas, M. (2018). The social constitution of agency and responsibility. In K. Hutchison, C. Mackenzie, & M. Oshana (Eds.), *Social dimensions of moral responsibility* (pp. 110–136). Oxford University Press.

Vargas, M. (2021). Constitutive instrumentalism and the fragility of responsibility. *The Monist*, *104*(4), 427–442.

Velleman, J. D. (2000). *The possibility of practical reason*. Oxford University Press.

Velleman, J. D. (2001). The genesis of shame. *Philosophy and Public Affairs*, *30*(1), 27–52.

Vihvelin, K. (2013). *Causes, laws, and free will: Why determinism doesn't matter*. Oxford University Press.

Walker, M. U. (2002). *Moral contexts*. Rowman & Littlefield.

Wallace, R. J. (1994). *Responsibility and the moral sentiments*. Harvard University Press.

Wallace, R. J. (2011). Dispassionate opprobrium: On blame and the reactive sentiments. In R. J. Wallace, R. Kumar, & S. Freeman (Eds.), *Reasons and recognition: Essays on the philosophy of T. M. Scanlon* (pp. 348–372). Oxford University Press.

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

Watson, G. (2004). Responsibility and the limits of evil: Variations on a Strawsonian theme. In *Agency and answerability* (pp. 219–259). Clarendon Press.

Watson, G. (2011). The trouble with psychopaths. In R. J. Wallace, S. Freeman, & R. Kumar (Eds.), *Reasons and recognition: Essays on the philosophy of T. M. Scanlon* (pp. 307–331). Oxford University Press.

Watson, G. (2014). Peter Strawson on responsibility and sociality. In D. Shoemaker (Ed.), *Oxford studies in agency and responsibility* (Vol. 2, pp. 15–32). Oxford University Press.

Wedgwood, R. (2002). The aim of belief. *Philosophical Perspectives*, *16*, 267–297.

Westlund, A. C. (2003). Selflessness and responsibility for self: Is deference compatible with autonomy? *The Philosophical Review*, *112*(4), 483–523.

Williams, B. (1995). Internal reason and the obscurity of blame. In B. Williams (Ed.), *Making sense of humanity and other philosophical papers* (pp. 35–45). Cambridge University Press.

Wolf, S. (1990). *Freedom within reason*. Oxford University Press.

Wolf, S. (2011). Blame, Italian style. In R. J. Wallace, R. Kumar, & S. Freeman (Eds.), *Reasons and recognition: Essays on the philosophy of T. M. Scanlon* (pp. 332–347). Oxford University Press.

Wright, L. (1973). Functions. *The Philosophical Review*, *82*(2), 139–168.

Wright, L. (1976). *Teleological explanations: An etiological analysis of goals and functions*. University of California Press.

Zimmerman, M. J. (1988). *An essay on moral responsibility*. Roman & Littlefield.

Zimmerman, M. J. (2010). Responsibility, reaction, and value. *Journal of Ethics*, *14*(2), 103–115.