

The Problem of Blame

Part of what makes the problem of blame so difficult is the fact that there is little consensus on the nature of blame itself. For example, some take blame to be a mere judgment with no necessary emotional sting accompanying it. Others take blame to be essentially accompanied by some angry emotional charge. The varieties of blame come packaged with very different accounts of the *sting* of blame, and thus the corresponding degree of discomfort or harm that blame causes differs drastically depending on the details.

In what follows, I will be restricting my focus almost exclusively to one particular variety of blame: *reactive blame*. The motivation for this restriction is that it is the kind of blame that motivates blame curmudgeons' skeptical worries about permissibility. As noted in the Introduction, I am assuming a stipulative characterization of curmudgeonly views about blame. While there are now a wide array of descriptively skeptical views about free will, moral responsibility, and blame (see, for example, Ishtiyaque Haji (2016), Neil Levy (2011), and Galen Strawson (1993)), curmudgeons are those who argue further for some prescriptive variety of eliminativism whereby we ought to eliminate reactive blame from our lives insofar as we can. The paradigm curmudgeonly views that I have in mind as targets here are Pereboom (2001, 2013), Caruso (2012, 2015, 2020), Strawson (1994), and Waller (1990, 2011, 2015, 2020).

Curmudgeonly worries do not arise for *every* kind of blame. It is not at all clear, for example, why one might be concerned about the permissibility of merely holding the belief that a wrongdoer is blameworthy. But, matters are different when we consider varieties of blame that are clearly harmful. The idea that we ought to avoid harming others without justification is among the bedrock of our moral intuitions, and so as we consider more harmful varieties of blame, the justificatory stakes creep higher, and concerns about permissibility naturally begin to emerge. And reactive

blame appears to be the *most* harmful variety of blame.¹ It is the kind of blame that is intimately tied to one of our most unpleasant moral emotions – moral anger.

Some attempts to defend blame from curmudgeonly skepticism have adopted the strategy of *avoiding* this kind of blame, arguing that there are less harmful, *sanitized* varieties that can sustain the valuable features of our blaming practices without the emotional costs of moral anger.² While there are merits to this kind of approach to defending blame against skepticism – for example, it carves out conceptual space for salvaging at least *some* variety of blame if such skeptical worries ultimately prove decisive – one might worry that this kind of sanitizing approach gives far too much away to skeptics from the start. For my part, I think that this strategy avoids the problem of blame by merely sidestepping it, leaving the real *hard problem* of blame largely untouched. No one – including the staunchest of blame curmudgeons – is losing much sleep over the permissibility of sanitized, painless blame. It is the nasty, reactive, angry kind that is controversial, and so the real challenge of defending blame against skeptical worries is to defend the reactive variety in particular.

My goal throughout this book is to meet this challenge head-on, and offer a defense of reactive blame in particular. If this defense succeeds, then it is not at all clear what further worries about the permissibility of blame could remain. If a case can be made for thinking that even the *worst* kind of blame is sometimes permissible, and that our all things considered reasons count in favor of retaining this kind of blame, then the most serious skeptical threats to blame will have been defused.

But what counts as reactive blame, and how ought we to distinguish this kind of blame from other more innocuous varieties? In Section 1.1, I trace a fault line between reactive and nonreactive varieties of blame and identify some of the features that unify reactive accounts. In doing so, I attempt to clarify a feature of reactive blame that has thus far led to a great deal of confusion – how we ought to understand the kind of *central role* played by the reactive attitudes for such accounts. I identify and set aside one particularly implausible (yet seemingly widespread) way of understanding this role, and identify two far more plausible alternatives.

¹ At least among plausible contemporary accounts. Here, I set aside accounts tied directly to retributive theological views or theories of punishment.

² Scanlon's (2008) conative view of blame is perhaps the most notable example. I discuss this view in further detail in Section 1.2.2.

In Section 1.2, I offer a rough taxonomy of explicitly nonreactive varieties of blame. Here I focus on cognitive and conative accounts as nonreactive paradigms. In Section 1.3, I examine whether certain accounts of blame that are more difficult to categorize – quality of will accounts and functionalist accounts – ought to be characterized as reactive or nonreactive. I conclude that the negative reactive attitudes play a sufficiently central role for the majority of these accounts to give rise to the problem of blame, and thus they are best characterized as reactive.

Finally, in Section 1.4, I turn to explicating the problem of blame itself. Once we recognize that the most vexing skeptical challenges to blame concern reactive blame and its attendant harms, a helpful analogy to a more familiar problem arises. This more familiar problem is *the problem of punishment*, and I argue that attempts to resolve this problem are instructive for understanding the problem of blame. In particular, attempts to resolve the problem of punishment help to explicate two clear desiderata for resolving the problem of blame, desiderata that if met should satisfy even the most skeptical blame curmudgeon. I devote the remainder of Part I to arguing that both of these desiderata can ultimately be met by a plausible account of reactive blame.

1.1 Reactive Blame

As noted earlier, part of the difficulty in articulating the problem of blame is the fact that there is little consensus about what the target concept actually *is*. Furthermore, there are significant methodological disagreements about how we ought to approach theorizing about blame in the first place. While some adopt a method akin to traditional conceptual analysis, others take paradigm instances of blame to be the best starting point, while still others think that we should begin by analyzing the work we want blame to do for us in our moral lives and offer an account that can make sense of blame's functional role.

For now, I will largely set these methodological differences aside, though I will return to them in Part II. In the next three sections, my goal is only to provide a rough taxonomy of the varieties of blame currently on offer, and a clearer method for distinguishing reactive blame from its less harmful counterparts. To that end, I will carve up the terrain along one central fault line: accounts of blame that take the negative reactive attitudes to play some central role, and those that do not. Those that fall under the former category are the varieties of blame I will call *reactive blame*, the kind of blame that the hard problem of blame is actually a problem for.

I will have a great deal to say about the positive components of reactive blame in the following chapters, and so wish to avoid any substantial commitments just yet. A full blown account of reactive blame (or, for that matter, any account of blame) should tell us something about the *object* of blame (what blame responds to), the *content* of blame (what, if any, propositional content is involved in blaming), what kinds of *responses* count as genuine blaming responses (does blame require some kind of communicative expression, or does private blame count?), and what the *aims* of blame are (what kinds of valuable or evolutionarily useful things does blame do for creatures like us?). I will discuss all of these features of reactive blame in substantial detail in Chapters 2–4. Here only a few preliminary remarks are in order for the task of initially distinguishing reactive blame from nonreactive varieties more generally.

First, it is important to note that I will be casting the net of reactive blame quite broadly. What I call reactive accounts will encompass a wide range of views sometimes characterized as *Strawsonian* in spirit, due to the central role played by the negative reactive attitudes that feature so prominently in P. F. Strawson's (1962) *Freedom and Resentment*. While nearly every other feature of blame may differ across reactive accounts of blame, what they all have in common is some kind of central role for these affective attitudes. But which attitudes? This question will feature prominently in the discussion of blame and the negative reactive attitudes in Chapter 3, but for now the reader can assume *at minimum* the relevant attitudes will be the varieties of moral anger: resentment, indignation, and guilt.

What do I mean when I say that these moral emotions play a *central* role in reactive accounts of blame? I take confusion about this question to have motivated a significant amount of criticism directed toward reactive accounts thus far. So, for the remainder of this section, I will attempt to clarify three possible answers. While I take the first answer to be the most widespread, I think it is also the least charitable way of construing the kind of centrality that is granted to the negative reactive attitudes for reactive blame. I also take this construal to give rise to much of the criticism just mentioned. As such, I will offer two alternatives that I think proponents of reactive accounts of blame can and should endorse instead.

1.1.1 *Reactive Essentialism*

What all reactive accounts of blame have in common is that the negative reactive attitudes play *some* central role in making sense of what blame is.

But the way that we understand the degree of centrality can make a significant difference when it comes to the plausibility of a particular reactive account of blame. Here I will characterize three main options for understanding this role that have coalesced in the literature thus far: (1) reactive essentialism, (2) functional reactivity, and (3) canonical reactivity.

The first option, which I will call *reactive essentialism*, is perhaps the most natural way of understanding reactive blame. It is also the one that opponents of reactive blame have called the most attention to. Reactive essentialism is the view that some experience or expression of the negative reactive attitudes is a *necessary* feature of blame. Thus, an instance of apparent blame will count as genuine only if it is accompanied or characterized by the actual experience or expression of a negative reactive attitude.

I take this way of understanding reactive blame to be something of a straw man. While much ink has been spilled arguing that reactive blame so understood is descriptively implausible – it is subject to a wide array of counterexamples³ – it is not at all clear that any proponents of reactive blame actually endorse reactive essentialism.⁴ In fact, many of those who have defended accounts of explicitly reactive blame most vigorously have themselves acknowledged such counterexamples and weakened their views accordingly. R. Jay Wallace, for example, argues only that blame involves a *susceptibility* to the reactive emotions,⁵ stating explicitly that “it is not required that we actually feel the relevant emotion in all the cases in which it would be appropriate to do so” (Wallace, 1994: 77). Wallace instead characterizes the connection between the reactive attitudes and blame as indirect and disjunctive. Blame is about holding one another to expectations we accept, and the reactive attitudes are constitutively linked to these expectations:

... to hold someone to an expectation, I suggest, is to be susceptible to a certain range of emotions if the expectation is violated, or to believe that it would be appropriate for one to feel those emotions if the expectation is violated. (Wallace, 1994: 23)

I will discuss Wallace’s view in much greater depth in Chapter 3, but here I wish to emphasize that even for Wallace – whose view is often taken to be a paradigm reactive account – the role played by the reactive attitudes is not an essential one. While susceptibility to the reactive attitudes is a *sufficient* condition for blame on Wallace’s view, it is not a necessary one.

³ See for example Sher (2006) and Scanlon (2008).

⁴ Susan Wolf (2011) is perhaps one notable exception. ⁵ See Wallace (1994: 12).

Rather, one might also hold someone to an expectation and blame them by meeting the second disjunct above, and simply believing that it *would* be appropriate for one to feel the reactive attitudes in a given circumstance.⁶

Wallace is not the only proponent of reactive blame to avoid commitment to reactive essentialism.⁷ In fact, this kind of reactive essentialism is quite difficult to find clear examples of.⁸ Nor do I see any clear reason for a proponent of reactive blame to accept reactive essentialism. There are alternative ways of understanding the central role that these attitudes play which are far more plausible. In what follows, I will set reactive essentialism aside, and focus instead on characterizing reactive blame in terms of one or both of these two alternatives.

1.1.2 *Functional Reactivity*

Angela Smith (2013) has recently offered an alternative to reactive essentialism, which I will here call *functional reactivity*. Smith is primarily interested in the role that moral protest plays in blame, arguing that registering one's moral protest is an essential feature of blame left out by many nonreactive accounts. She attempts to carve out a middle ground between "moral assessment" accounts of blame, and "moral sanction" accounts, taking blame to involve more than a mere negative judgment about its target's quality of will while still falling short of full blown punishing or sanctioning (Smith, 2013: 27).

Smith states the view as follows:

To morally blame another, in my view, is to register in some significant way one's moral protest of that agent's treatment of oneself or others. (Smith, 2013: 29)

⁶ Here my claim that Wallace does not endorse reactive essentialism might be puzzling to some, given Wallace's well-known criticism of nonreactive accounts like Scanlon's for "leaving the blame out of blame" (Wallace, 2011: 349). Wallace's criticism targets the fact that, on Scanlon's account, the negative reactive attitudes are at best only contingently related to blame. We could fully account for genuine blame on Scanlon's view without them. On Wallace's view we could not, though this need not entail the kind of strict reactive essentialism discussed here. As I will discuss later, Wallace and others can make sense of the centrality of the negative reactive attitudes without going so far as to claim that their experience or expression is a necessary condition for genuine instances of blame by appealing instead to *functional* or *canonical* reactivity.

⁷ Even Strawson himself makes no explicit commitment to *this* degree of centrality for the reactive attitudes, despite the important role he takes these attitudes to play in constituting our blaming and responsibility-related practices more broadly.

⁸ As noted earlier, Wolf might be one exception and goes so far as to say that "liability to feel angry emotions and to form angry attitudes appears to be an inevitable feature of allowing oneself to be not just physically but emotionally vulnerable to other people" (Wolf, 2011: 337).

While registering such protest “need not take the form of a Strawsonian reactive attitude,” Smith acknowledges that explicitly Strawsonian accounts do *come closest* to capturing the element of moral protest that she does take to be essential. Understood as a form of moral protest, Smith then takes the function of blame to be twofold. First, blame registers the fact that the victim did not deserve a certain kind of treatment by “*challenging* the moral claim implicit in the wrongdoer’s action” (Smith, 2013: 43). Second, blame “prompts moral recognition and acknowledgement of this fact on the part of the wrongdoer and/or others in the moral community” (Smith, 2013: 43). And resentment and indignation are *very common* ways of protesting the conduct of others and meeting these two aims. However, these negative reactive attitudes are not the *only* means of doing so. It is possible, for example, to challenge and prompt recognition of such conduct “dispassionately,” sometimes by merely modifying our attitudes, intentions, and expectations of one another (Smith, 2013: 45).

The fact that the negative reactive attitudes are not a necessary feature of moral protest rules out reactive essentialism for Smith. However, Smith’s account of blame still has a significantly reactive feel, and she ultimately concludes:

Of all of the traditional and contemporary accounts of blame on offer, it seems to me that the Strawsonian account comes closest to capturing this crucial aspect of these distinctively moral responses, which perhaps explains why his view has had such staying power. To the extent that it fails, it is only in placing too much emphasis on just one – albeit one very important – set of emotional reactions as the *sine qua non* of moral protest. (Smith, 2013: 48)

Here an alternative to reactive essentialism begins to emerge. Rather than taking the negative reactive attitudes as an *essential* feature of reactive blame, we might more charitably understand the centrality of the negative reactive attitudes to blame in something like Smith’s terms. While these attitudes are not strictly necessary, they are often the *best*, or at least *the most common* means we have of meeting and sustaining the aims of blame. Call this understanding of the way in which the negative reactive attitudes are central to reactive blame *functional reactivity*.

On Smith’s view, the aims of blame directly involve moral protest, but one need not be committed to this detail in order to embrace functional reactivity. Here I simply take Smith’s view to be instructive in elucidating a plausible alternative to reactive essentialism.

1.1.3 Canonical Reactivity

Functional reactivity is not the only alternative to reactive essentialism. Victoria McGeer has also recently applied Jackson and Pettit’s (1995)

appeal to canonical features of a kind to blame, and in doing so, sketched an additional alternative. Canonical features are those features that have criterial significance, in that they “account for our interests in identifying a kind as such, even though things belonging to the kind do not invariably manifest the feature in question” (McGeer, 2013: 168). Canonical features are more than mere characteristic or typical features of a kind, but also fall short from counting as essential or necessary. McGeer suggests that the negative reactive attitudes are precisely this kind of feature of blame. I will call McGeer’s method of characterizing the centrality of the negative reactive attitudes to blame *canonical reactivity*.

This proposal for understanding the centrality of the negative reactive attitudes arises in the context of McGeer’s attempt to resolve a tension similar to the one at issue for what I have been calling the hard problem of blame – blame seems to be an important and valuable way of responding to wrongdoing, yet often has an angry, punitive edge that seems difficult to justify. McGeer argues that the project of resolving this tension and thus “civilizing blame” is subject to two constraints. First, she finds attempts to simply abandon reactive, angry blame problematic, arguing that a *psychologically realistic* account of blame must be an account of the reactive variety, “taking blame to be a phenomenon that displays the negative profile (warts and all) that is typically associated with [it]” (McGeer, 2013: 163).⁹ Further, any attempt to civilize blame can only meet a second *normative constraint* by explaining precisely how this “unsavory” kind of blame “can still do valuable normative work so long as it is constrained by social and institutional practices that support its more constructive features” (McGeer, 2013: 163). For McGeer, it is precisely the negative reactive attitudes, understood as a canonical feature of blame, which do this valuable normative work.

I will discuss McGeer’s view in further detail in Chapter 4, and here I wish to focus only on her characterization of the negative reactive attitudes as canonical features of blame. This characterization presents yet another plausible way of understanding the central role that the negative reactive attitudes might play for reactive blame while avoiding full blown reactive essentialism. McGeer’s account is similar to the one sketched by Smith in that it takes the functional role of blame to be an important starting point. However, her own view of the functional role of blame differs in that McGeer is primarily concerned with blame as a

⁹ As discussed earlier, I am in significant agreement with McGeer on this point.

psychological kind, taking a naturalistic, evolutionary approach to understanding its aims.

According to McGeer, blame involves both backward and forward-looking elements of appraisal. In regard to the former, blame is triggered by various events only insofar as they are coded or perceived as situations of a certain emotion-inducing type, namely as norm transgressions (McGeer, 2013: 171). In regard to the forward-looking element, blame disposes us to behave in ways that at one point had an overall tendency to enhance individual fitness in ancestral populations by promoting large-scale human cooperation (McGeer, 2013: 171). On McGeer's account, the psychological feature that plays both these forward- and backward-looking roles is *moral anger*. Moral anger is the psychological state that *typically* plays the causal role of blame in creatures like us, and the unpleasant emotional character of this state is *not incidental* (McGeer, 2013: 169). While it is possible to blame dispassionately without moral anger (and thus reactive essentialism is false), such instances of blame are in some sense *defective* on McGeer's account. They lack a *canonical feature* of blame, namely the negative reactive attitudes that constitute the varieties of moral anger.

But how ought we to understand the distinction between canonical features, and other kinds of features? Here McGeer is also instructive. She distinguishes between features of a kind that are essential or constitutive, those that are merely "kind-associated," and those that are canonical. Essential features are, again, those necessary to a genuine instance of the kind, whereas mere kind-associated features are "features whose contingent association with the phenomenon in question is of no criterial significance" (McGeer, 2013: 168). The ability to fly, for example, is a kind-associated feature of birds. While many members of the kind have this feature, it does not figure in how we demarcate the bounds of the kind itself. The ability to fly is a contingent feature of birds, and not the sort of thing we appeal to in order to distinguish birds from non-birds, or to understand the kind of thing birds are.¹⁰

Canonical features, in contrast, are those that *do* have criterial significance. They are the very features that

... account for our interest in identifying the kind as such, even though things belonging to the kind do not invariably manifest the feature in question. (McGeer, 2013: 168)

¹⁰ Perhaps Smith's account might best be characterized as one for which the negative reactive attitudes are merely "kind-associated," while the fact that they are the *best* kind-associated features for sustaining the functional role of blame is what renders this account a functionally reactive one. Thanks to Neal Tognazzini for suggesting this point.

Canonical features therefore represent a middle ground between mere kind-associated features and features that are essential. Genuine members of the kind can lack a canonical feature, but we could not make sense of the kind itself without such features. Here McGeer's use of the kind "heart" as an example is instructive:

Whether natural or artificial, hearts are canonically the kind of things that pump blood. A heart deserves to be called a heart only so far as it is a thing of *that* (instrumental) kind. But, of course, this doesn't mean that all hearts are able to pump blood. Things of that kind sometimes malfunction. (McGeer, 2013: 168)

Just as hearts are canonically the kinds of things that pump blood, blame is canonically the kind of thing that involves moral anger. Yet, just as a heart that has stopped beating is still heart, so too might a dispassionate instance of blame still count as genuine blame. However, like the stopped heart, such an instance of blame would be defective in some sense. Blame is *the kind of thing* that usually involves moral anger, and this negative reactive attitude serves as a canonical feature.

I take both functional reactivity and canonical reactivity to represent plausible ways of understanding the way in which the negative reactive attitudes are central to reactive blame. Here I will not take a stand on which of these alternatives to reactive essentialism I am inclined to endorse. Nor is it clear that these two characterizations should necessarily be treated as distinct options. It is worth emphasizing that the appeal to the functional role of blame is an important feature of each way of understanding the centrality of the negative reactive attitudes for reactive blame, and that one might easily combine a normative approach to understanding this role like Smith's with a naturalistic, evolutionary approach to understanding it like McGeer's.¹¹

In what follows, I assume only that reactive essentialism is an uncharitable way of characterizing reactive accounts of blame. Critics are correct to point out that counterexamples involving apparently genuine yet dispassionate instances of blame abound, but incorrect to dismiss reactive blame because of them. There are other plausible ways of construing the central role of the negative reactive attitudes for reactive blame that need not characterize these emotions as necessary or essential. In what follows,

¹¹ In fact, this will largely be my approach in Chapter 4.

I invite the reader to think of the role of these attitudes, especially moral anger, in terms of either *functional reactivity*, or *canonical reactivity*.

I will use these two ways of thinking about centrality to distinguish reactive accounts of blame from their nonreactive counterparts. As we will see in Section 1.3, it is not always clear whether an account of blame is nonreactive or reactive to a degree relevant to the problem of blame. And this is especially true once we have abandoned reactive essentialism as a way of characterizing the latter. In what follows, I will use the disjunction of functional and canonical reactivity as a test for the degree of reactivity needed to give rise to the problem of blame. If an account takes the negative reactive attitudes to be necessary to explaining or sustaining the functional role of blame, or to distinguishing blame as a genuine kind in the first place, then it is an account of *reactive blame* in need of defense from the problem of blame.

I turn now to a brief tour through two explicitly nonreactive alternatives to reactive blame, accounts that seem to easily avoid the problem of blame and worries about permissibility.

1.2 The Varieties of Nonreactive Blame

Work on blame has thus far occurred across a wide range of philosophical subdisciplines, and as such has not yet lent itself to anything like a canonical taxonomy. One noteworthy exception in recent work has been Justin Coates and Neal Tognazzini's (2013) clear and informative attempt to carve out the various areas of the terrain in their introduction to *Blame: Its Nature and Norms*. In this section, I will largely borrow from their chosen method of categorization, with a few noteworthy departures that I think help to better sharpen the distinction between reactive and non-reactive blame in particular.

Coates and Tognazzini carve up the terrain of blame in terms of four main categories: cognitive accounts, conative accounts, Strawsonian accounts, and functional accounts. Strawsonian accounts correspond straightforwardly to what I have here been calling reactive blame, as accounts that explicitly take the negative reactive attitudes to play some central role. Here I will set Strawsonian accounts aside and focus on the two main categories of paradigmatically nonreactive blame: cognitive accounts and conative accounts. I will then turn in Section 1.3 to discussion of accounts that are a bit more difficult to characterize as reactive or nonreactive, and offer some suggestions for how we ought to make this distinction in light of the discussion of functional and canonical reactivity.

1.2.1 *Cognitive Accounts of Blame*

If reactive blame occupies one extreme on the spectrum of the potential harms of blame, *cognitive accounts* lie at the other. Cognitive accounts are those that take blame to be some kind of *evaluative judgment*, where the content of this judgment varies. For example, some cognitive accounts take the content of the evaluative judgment involved in blame to be a kind of negative mark on one's "moral ledger" (Zimmerman, 1988: 38).¹² This kind of view was proposed by J. J. C. Smart (1961), who characterized blame as similar to dispraise in that it involves a *grading* judgment of its target as a thing of a certain kind.¹³ However, for Smart blame goes beyond mere dispraise in that this evaluative judgment has some additional significance. It implies that its object is responsible for their action or character.

Other cognitive accounts take blame to involve an aretaic judgment, or as Gary Watson puts it, to see the conduct at issue "as a poor exercise of human evaluative capacities, as characteristic of someone who cares little about standards of excellence in human affairs" (Watson, 2004: 265). While Watson does not take this kind of cognitive blame – the kind associated with what he calls the *attributability* face of responsibility – to be the only kind of blame, it is itself a robust cognitive alternative to reactive blame.

What the variety of cognitive accounts have in common is that *some* evaluative judgment is not only necessary, but also sufficient for blame. Further, these accounts make no mention of the reactive attitudes. The varieties of moral anger play *no* role in blame on a straightforwardly cognitive account, though they may turn out to be merely contingently associated with the relevant evaluative judgment. Thus, cognitive accounts constitute perhaps the clearest example of straightforwardly nonreactive blame.

However, not all evaluative judgments are normatively equal. This is especially clear when we consider evaluative judgments about one's *quality of will*. Judgments about quality of will seem to have a special normative force, and are importantly distinct from other kinds of negative evaluative judgments, say about one's awful cooking or bad driving. Quality of will *matters more to us*, and judgments about it flag ways in which someone has "dropped below some standard that we accept (or perhaps that we think

¹² For similar views, we might classify as "ledger" views; see Glover (1970) and Haji (1998).

¹³ See also Schlick (1939).

they should accept), whether of excellence, morality, or respectful relationships” (Coates and Tognazzini, 2013: 9). We care about such judgments in ways that we simply do not care about what others think about our cooking or our driving, even (and perhaps *especially!*) if they are never communicated to us.

Views that take blame to be constituted by an evaluative judgment of ill will are often categorized as cognitive accounts. However, for my current purposes things are not so clear cut. Because judgments of ill will have special significance and force for us, they are precisely the kind of judgments that tend to trigger the negative reactive attitudes. Whether or not such views end up counting as reactive or nonreactive will depend on the details (namely whether or not the negative reactive attitudes turn out to be an essential, functionally necessary, or canonical feature of blame), and as such I will treat them separately in Section 1.3. Before dealing with these thornier issues, it will be first helpful to get a clear picture of the other paradigmatic variety of nonreactive blame, conative blame.

1.2.2 Conative Accounts of Blame

In addition to cognitive accounts, *conative accounts* represent perhaps the most fully developed and plausible alternative to reactive blame. Among the most prominent defenders of conative accounts of blame are George Sher (2006) and T. M. Scanlon (2008). Conative accounts follow cognitive accounts in taking some kind of evaluative belief or judgment to be central to blame, but depart from these more minimal accounts in that the relevant belief is not sufficient. Conative accounts take some corresponding change to the blamer’s motivation – their desires, attitudes, or other dispositions to behave in certain ways – to also be necessary. Here I will focus on Sher and Scanlon’s accounts as the two most fully developed conative accounts currently on offer.

According to Sher’s conative account, blame is a belief–desire pair. The belief in question is that the target of blame has acted badly or has a bad character, while the corresponding desire is the desire that the target *had not* acted badly, or *had not* had a bad character (Sher, 2006: 112). In regard to the belief component, Sher clarifies that “bad acts” refer to “morally defective acts that render agents blameworthy” (Sher, 2006: 9). So, the belief component looks to be simply a judgment that the target of blame has a character or has committed some wrong conduct for which they are blameworthy. It is then the desire component of the core pair that allows this account to avoid vicious circularity by further elucidating what the

blameworthy are worthy *of*— namely, whatever reactions are made appropriate by the overall pair.

This core belief–desire pair composes the first “tier” of Sher’s account of blame, with some class of blame-related behavioral and attitudinal dispositions forming the second tier (Sher, 2006: 138). According to Sher, it is the desire component of the core belief–desire pair that explains the relevant behavioral and attitudinal dispositions:

[T]he obvious way to invoke D [the desire that the target of blame had not had a bad character or acted badly which makes up the desire component of the belief–desire core of blame] to account for our disposition to become angry at those we blame is to assimilate that anger to the other negative feelings that we have when we see that we cannot get what we want. Just as obviously, the way to invoke D to account for our disposition to display hostility toward those we blame is to see our hostile behavior as a natural expression of our negative feelings toward them. (Sher, 2006: 104–105)

Sher’s key insight is that there is something motivational at the heart of our blaming reactions. His view has the advantages of allowing for a significant degree of variation across our actual blaming responses and explaining why the negative reactive attitudes are not necessary to blame.¹⁴ In regard to the latter, consider a particularly close relationship with a friend or loved one. In this case, the belief that she is blameworthy paired with a desire that she had not wronged you might dispose you to other responses than hostile negative reactive attitudes. Hoping to maintain the relationship you might instead be disposed to calmly confront her, and communicate information about the kind of amends or apology that would be needed to reconcile fully. Many have criticized Sher’s claim that the core belief–desire pair he identifies is sufficient to explain a wide enough range of apparent instances of blame, or even that it is necessary to do so.¹⁵ Here I do not wish to assess the overall plausibility of Sher’s conative view of blame, but merely to note it as a robust conative alternative to reactive blame.

Perhaps, the most richly developed and defended alternative to reactive blame is another conative account, the account developed over the past several decades by T. M. Scanlon (1988, 1998, 2008). Scanlon (2008) explicitly proposes an account of blame that he takes to satisfy four main desiderata, namely to explain: (1) the difference between blame and mere

¹⁴ This core belief–desire pair remains constant across all agents and contexts, but will make different reactions appropriate depending on the relationship between the blamer, wrongdoer, and victim.

¹⁵ For persuasive arguments to this end, see Smith (2013).

objective stigma or *moral outcome luck*; (2) the relation between blame and *wrongness*; (3) the *significance* of blame and why blame matters to us; and (4) the *ethics* of blame (who can be the target of blame, who has standing to blame, and why we should blame).¹⁶

According to Scanlon, to *be blameworthy* entails that something about your attitudes toward others *impairs* the relationship that they can have with you. This notion of impairment is central to Scanlon's account of blame, and it arises out of his account of the standards constitutive of our interpersonal relationships (Scanlon, 2008: 131–138). These relationships are constituted by the *reasons* one takes themselves to have for treating others in certain ways, and corresponding *dispositions* to feel and act in certain ways. Nondefective interpersonal relationships should be mutual or reciprocal and governed by certain normative standards and ideals. The relevant kind of normative ideal is simply whatever must be true in order for an individual to have a relationship of the relevant kind; how they should behave ideally; and what attitudes they should have ideally relevant to that kind.

On Scanlon's view, different kinds of relationships can, and in fact sometimes do, have radically different normative ideals. The conditions that must be met in order for me to be a term in an interpersonal romantic relationship will be drastically different from the kinds of ideal attitudes, dispositions, and expectations needed to engage in a relationship between colleagues. These normative ideals set the standards for a kind of relationship, and also restrict what might count as a genuine interpersonal relationship of the relevant kind. While on Scanlon's view there will be a wide array of kinds here, not just any relation will count. One-off or grue-like¹⁷ relations (say between people who see each other in the hallway in passing on Tuesdays) will not count, because there is no apparent normative ideal relevant to them. Finally, on Scanlon's view most of our interpersonal relationships are *contingent* and *conditional*.¹⁸ Maintaining them often depends on several factors that can lie outside of one party's control, and

¹⁶ See Scanlon (2008), chapter 4.

¹⁷ "Grue-like" relations, like grue-like predicates, would be those that are very poorly *entrenched*. In the past, we have not projected many hypotheses involving them or relations coextensive to them. In other words, they are the kind of relations that we do not often (or ever) find ourselves in need of in order to make useful explanatory or inductive claims, and they are likely to be massively or randomly disjunctive. They are relations that do not "hang together" in any apparent way. For further discussion see Goodman's (1955) account of grue-like predicates.

¹⁸ One notable exception is *the moral relationship* that Scanlon posits between all moral-reasons-responsive agents in order to avoid the objection that his view cannot make sense of blaming strangers. See Scanlon (2008: 139–152).

they can be *ended*. Often relationships come to end due to an impairment. But they can also be ended blamelessly and without impairment, such as when two friends simply drift apart.

Like blameworthiness, actively blaming someone is closely related to the normative ideals of our interpersonal relationships and what it means to impair them. On Scanlon's view, an interpersonal relationship is *impaired* when one party, while standing in the relevant relation to another person, holds attitudes toward that person that are ruled out by the standards (normative ideals) of that relationship, thus making it appropriate for the other party to have attitudes other than those that the relationship normally involves. This notion of impairment is not all or nothing – most of our interpersonal relationships can continue in an impaired form. A bad friend, for example, might still be considered a genuine friend (Scanlon, 2008: 131–138). When a relationship has been impaired in the relevant sense, it is appropriate *to blame* the party who has caused the impairment. Scanlon explicitly defines what it means to blame someone as having two features: (1) a judgment that the target of blame is *blameworthy* (as stated earlier, the belief that they have impaired the relevant relationship); and (2) to take the relationship with the target of blame to be *modified* in a way that this judgment renders appropriate.

Here the notion of *modification* still needs to be elucidated, and it is this feature that renders Scanlon's account of blame a conative one. Scanlon uses the example of Joe, a disappointing friend, to articulate some of our options for modification:

Suppose I learn that at a party last week some acquaintances were talking about me, and making some cruel jokes at my expense. I further learnt that my close friend Joe was that party, and that rather than coming to my defense or adopting a stony silence, he was laughing heartily and even contributed a few barbs, revealing some embarrassing facts about me that I had told him in confidence. This raises some question about my relationship with Joe. . . . Possible responses, on my part, to what Joe has done fall into three general categories. First, I might consider whether I should continue to regard Joe as a friend. An answer to this question is a judgment about the meaning of Joe's action – about what it shows about his attitude toward me, considered in relation to the requirements of friendship, and about the significance of that attitude for our relationship. Second, I might revise my attitude toward Joe in the way that this judgment holds appropriate. I might, for example, cease to value spending time with him in the way one does with a friend, and I might revise my intentions to confide in him and to encourage him to confide in me. Third, I might complain to Joe about his conduct, demand an explanation or justification, or indicate in some other way that I no longer see him as a friend. (Scanlon, 2008: 129–130)

While Scanlon takes it to be a merit of his account that it allows for a wide array of relationship-modifying responses to count as blaming responses, this example highlights a few noteworthy possibilities. First, and perhaps most minimally, modification might involve simply altering your judgments and attitudes about the nature of your relationship with the target of blame. A betrayal like Joe's might lead you to view him as more of an acquaintance than a friend. In doing so, you might change your intentions toward Joe, and adjust your expectations of him in the future. For example, you may no longer intend to spend substantial amounts of time with him, or to confide in him. Or, you may abandon the expectation that he will keep any future confidences, or defend you against slander from others.

Whatever the attitudinal revision, this first kind of modification need not entail any overt behavior or speech that communicates the modification of your relationship. If the target of blame is not particularly observant, then it is compatible with Scanlon's view that this kind of blaming modification might occur without any uptake from that target. In practice, it does seem that blame is sometimes "silent" in this way, and so one virtue of Scanlon's account is that it has the resources to make sense of this kind of *private* blame.

But blame is often public, and does in fact aim at some form of uptake from its target. On Scanlon's account revising not only our attitudes but also our *dispositions* can mark a modification to a relationship to both the blamer and target of blame. Returning to Joe, my previous disposition to feel happiness when good things happen to Joe may weaken. Likewise, I might adjust or completely abandon my disposition to defend him when others speak poorly of him, to prioritize spending time with him, or to help him when he is in need. It is not clear how fine grained we ought to be in characterizing these kinds of dispositions, but in Scanlonian terms we might think of many of them under the umbrella of a certain kind of *withdrawal*. Perhaps, one of the most widespread ways in which we blame others we judge to be blameworthy is to mark the impairment they have caused by taking a step back from them. We withdraw, and adjust many of the ways that we were previously disposed to engage with them.

This second kind of modification is far more likely to receive uptake and be noticed by blame's target. It is difficult for someone engaged in a genuine friendship not to notice when a friend withdraws or distances themselves. But it is still not the most overt way of marking the fact that an impairment has caused a particular relationship to be modified. Scanlon also acknowledges that we sometimes *express* our judgment that a

relationship has been impaired with an explicit complaint, demand for explanation, apology, or justification, or by pairing any of the above with an expression of resentment or indignation. However, this final possible means of modification has an entirely contingent relation to blame. While we *can* mark modifications to our relationships in response to a judgment of impairment in ways involving the reactive attitudes, we need not do so:

The account of blame that I offer is like Strawson's in seeing human relationships as the foundations of blame. But it differs from his view in placing emphasis on the expectations, intentions, and other attitudes that constitute these relationships rather than on moral emotions such as resentment and indignation. (Scanlon, 2008: 128)

On Scanlon's view, then, the negative reactive attitudes are neither an essential, functionally necessary, nor canonical feature of blame. As such Scanlon's account is explicitly nonreactive, and constitutes perhaps one of the most fully developed alternatives to reactive blame.

1.3 Quality of Will and Functional Accounts

While some accounts of blame are obviously reactive (for example, Strawson and Wallace's accounts), and others are clearly nonreactive (for example, Sher and Scanlon's accounts) some views are more difficult to place on either side of the fault line of reactivity. This is especially true of one subset of views often taken to be a kind of cognitive account – *quality of will accounts*. It is also true for accounts that take the functional role or aims of blame as their methodological starting point, a family of views that Coates and Tognazzini categorize as *functional accounts* of blame. Quality of will and functional accounts often attribute *some* important role to the negative reactive attitudes, but it is not always clear whether this role is central enough to give rise to the problem of blame.

While attempting to characterize all of the varieties of blame currently on offer along the fault line of reactivity is well beyond my current purposes, it may be helpful to treat at least one prominent version of each of these kinds of accounts as test cases. And with the distinction between reactive essentialism, functional reactivity, and canonical reactivity discussed in Section 1.1 now in hand, we have a clear method for doing so. In this section, I will take Pamela Hieronymi's (2004) quality of will account and Christopher Franklin's (2013) functional account as test cases, and assess the extent to which each of these views might be committed to reactive essentialism, functional reactivity, or canonical

reactivity. Ultimately, it seems that both views are committed to at least one of the latter, and I conclude the section with some brief remarks on whether we ought to draw any broader tentative conclusions about the reactivity of quality of will and functional accounts more generally (I think that we should).

1.3.1 *Quality of Will Accounts of Blame*

As discussed earlier, *quality of will accounts* of blame are sometimes categorized loosely as cognitive accounts. Like cognitive accounts, quality of will accounts take blame to involve a particular kind of evaluative judgment. However, unlike evaluative judgments about one's moral "ledger," or aretaic character, judgments about quality of will seem to matter to us in ways that feel distinctly *forceful*. And that is due in large part to the apparent connection between our beliefs about the quality of one's will and the negative reactive attitudes. The question, then, of characterizing such views as reactive or nonreactive will hang on what the view takes this connection to amount to. If the negative reactive attitudes are an *essential* feature of our evaluative judgments about one's quality of will, if they are necessary for such judgments to fulfill the *functional role* of blame, or if they are *canonical* features of such judgments, then the view in question will count as a reactive account of blame and one for which the problem of blame is in fact a problem. If not, then while the view might acknowledge that the negative reactive attitudes can (or even that they often contingently do) accompany our judgments about quality of will, then these views would be better characterized as nonreactive.

While there are a wide array of views that take an evaluative judgment about one's quality of will to feature significantly in blame,¹⁹ here I will focus on Pamela Hieronymi's (2004) influential account in particular. According to Hieronymi, blame is an evaluative judgment of ill will or disregard. Hieronymi's account is especially instructive for examining the centrality of the negative reactive attitudes for quality of will views, as she is explicitly interested in the *force* of blame, and whether or not such force could be subject to what she calls *the target charge of unfairness* (Hieronymi, 2004: 115).

The target charge of unfairness is the claim that the characteristic force of blame is unfair in certain conditions, namely when either: (1) the target

¹⁹ For a helpful taxonomy of the wide array of views of blame that take quality of will to be a central component, see Shoemaker (2015).

of blame could not have done otherwise and gotten themselves off the moral hook; (2) they did not have the ability to control their behavior in light of moral reasons; or (3) they did not play a sufficient role in becoming the kind of person that they are (Hieronymi, 2004: 116–118). Whether or not the target charge is true will depend on what the force of blame actually is, and on Hieronymi's view, the primary force of blame lies in the judgment of ill will that constitutes blame itself. Even though this judgment is descriptive "many truths are important to us," and it matters to us whether such judgments of ill will are true, or whether others think that they are true about us, because of the importance of our mutual regard for one another (Hieronymi, 2004: 122).

As noted in Section 1.2, judgments of ill will are important to us in ways that evaluative judgments about other features – for example, one's bad driving or cooking – are not. They are judgments about a kind of moral failure – their object has disregarded the worth or standing of a fellow person. According to Hieronymi, this is precisely what lends the evaluative judgment of ill will constitutive of blame its special force and significance. However, she argues that the target charge of unfairness does not apply to this particular kind of force, because the only thing that could render it unfair would be an absence of *object-focused reasons* for thinking that the judgment that gives rise to it is true.²⁰ According to Hieronymi, whether the target of blame could have done otherwise in a way that would get them off the moral hook or played a sufficient role in becoming the kind of person that they are is not relevant to whether an evaluative judgment that they have failed to regard others' moral worth or standing appropriately is fair. What matters is simply whether they have so failed. Barring any object-focused reasons to think that a judgment of ill will is false, we lack grounds for thinking that a corresponding instance of blame is unfair.

However, the special significance of evaluative judgments of ill will is not the *only* feature of blame that lends it its force that Hieronymi considers. She also notes that the disagreeable, adverse effects of the negative reactive attitudes that tend to correlate with such judgments sometimes lend blame some *further* force. However, according to Hieronymi the force of these attitudes is not located primarily in their affect, as the emotional aspect of these attitudes is largely involuntary.

²⁰ I discuss Hieronymi's distinction between object and attitude-focused reasons and how they might inform our understanding of what renders blame deserved in the basic sense further in Chapter 2. This distinction is also sometimes referred to using Parfit's (2001) terminology of "object-given" and "state-given" reasons, or in terms of epistemic versus pragmatic reasons (see also Reisner, 2009).

Instead, Hieronymi argues that the negative reactive attitudes lend blame an additional “multifaceted” force (Hieronymi, 2004: 133). While the affective component of these attitudes often serves to *mark* the significance of the evaluative judgment of ill regard at the heart of blame, their own force is itself derived largely from the significance of this judgment (Hieronymi, 2004: 133).

Whether or not Hieronymi’s account is best characterized as reactive or nonreactive will depend on the nature of this significance-marking role.²¹ Are the negative reactive attitudes *essential* to this role? Here the answer seems clearly no. Are the negative reactive attitudes *functionally necessary* to marking the significance of our judgments of ill will? This possibility seems far more plausible. Perhaps for creatures like us with the psychological makeup that we find ourselves with these unpleasant emotions are the only way to adequately mark this kind of significance.²² Finally, do the negative reactive attitudes seem to be a canonical feature of blame on this view? Here I am unsure, but a case could certainly be made for thinking that we might have difficulty understanding blame as the kind of thing that involves judgments of ill will that have the special force that they do without their relation to the negative reactive attitudes.

So, while Hieronymi’s account of blame is often characterized as a cognitive (and thus nonreactive) one, I think we have good reason to treat it as sufficiently reactive to give rise to worries about permissibility and the problem of blame. Hieronymi’s account of blame is likely to find itself in the curmudgeon’s crosshairs along with other more explicitly reactive accounts.

1.3.2 *Functional Accounts of Blame*

Another subset of views about blame that do not clearly lend themselves to the reactive or nonreactive distinction are *functional accounts*. Often such

²¹ This question is also distinct from how Hieronymi herself would characterize the view. Because she takes great care to locate the source of blame’s force in a set of judgments and not the reactive attitudes, it may be the case that she would push back on a reactive characterization in order to avoid breathing life back into the target charge of unfairness she so effectively defuses. However, my motivating reasons for getting a clearer picture of which accounts of blame we ought to consider reactive are different from Hieronymi’s. While Hieronymi is careful to move away from appeals to reactivity to explain the force of blame and thus how it might inform any potential charges of *unfairness*, I am primarily concerned with the all-things-considered permissibility of blame. We may very well have good reason (and in fact I will argue in Chapter 2 that we do have such reasons) to view these potential threats to blame as distinct. And so while I ultimately think Hieronymi’s view is sufficiently reactive to give rise to the problem of blame, this is not inconsistent with reasons to view it as nonreactive in distinct contexts where fairness is of primary concern.

²² Such an account would be very similar to McGeer’s (2013).

views have a *hybrid* feel to them, combining elements from cognitive, conative, and straightforwardly reactive accounts. What functional accounts have in common is the idea that *we should identify blame with its tasks*. Coates and Tognazzini characterize this kind of view in terms of the following methodology:

... we should figure out what *function* blame serves and then allow the particular context to determine which mental state or activity best serves that function, and so let context determine which way of responding counts as blame. (Coates and Tognazzini, 2013: 16)

Both Smith and McGeer's accounts might be characterized as functional accounts of blame, in that they both begin with an account of the aims of blame. Recall that for Smith, blame is a form of moral protest; and for McGeer, blame is *the kind of thing* triggered by perceived wrongdoing that in turns disposes creatures like us to behave in certain ways (McGeer, 2013: 171). As discussed in Section 1.1, I take Smith and McGeer's accounts of blame to each be instructive examples of plausible accounts of reactive blame that avoid commitment to reactive essentialism. In taking the role of the negative reactive attitudes to be the most paradigmatic means of achieving the aims of blame (in Smith's case), or to have criterial significance crucial to identifying the kind of thing that blame is (in McGeer's case), each view gives the negative reactive attitudes sufficient pride of place to give rise to the problem of blame curmudgeons are worried about. Will this be the case for other explicitly functionalist accounts of blame?

Whether a specific functionalist account of blame will best be characterized as reactive or not will ultimately depend on the details, in particular on how the aims of blame are characterized and the degree to which we need the negative reactive attitudes to sufficiently pursue or understand those aims. It may be helpful to note that I suspect most functionalist views will tend in the direction of reactivity, but here I will examine Christopher Franklin's (2013) *value account of blame* as a functionalist test case.

On Franklin's view, the aim of blame is to properly value the objects of moral value, and blame is "a mode of valuation required by the standards of value" (Franklin, 2013: 209). According to his preferred account of what it means *to value* something (as opposed to merely judging it valuable), valuing requires certain emotional and deliberative dispositions. Perhaps foremost among them is the disposition to protect and defend the objects one values. While Franklin avoids controversial commitments to

an account of what the proper objects of moral value actually are, he takes it as a given that they will include human beings (Franklin, 2013: 213–216). When such objects suffer an instance of moral wrongdoing, Franklin takes this to have a specific kind of symbolic meaning – wrong actions express the wrongdoer’s claim that the object wronged is not to be valued after all (Franklin, 2013: 218–223).

The role of blame in valuing the proper objects of moral value is to protect and defend these objects against such claims. And doing so requires more than mere expression of a judgment that the symbolic claim expressed by wrongdoing is false. Disagreeing is not defending, especially when it comes to claims as serious as the denial of one’s value as a person.²³ Instead, Franklin argues that properly valuing persons and truly defending and protecting them from such claims requires a disposition to blame wrongdoers in just the sort of way colored by the negative reactive attitudes. Mere judgments, or the kinds of dispassionate modifications to other dispositions sufficient to blame on conative views like Scanlon’s, will not be enough here. Franklin argues instead that what is required to fulfill the aims of blame and successfully protect and defend the proper objects of moral value when they are wronged is a disposition to experience and express the negative reactive attitudes (Franklin, 2013: 218–223).

In offering an account that takes dispositions to experience and express the negative reactive attitudes as crucial to achieving the aims of blame, it seems that Franklin’s account is best characterized as a reactive one. While the negative reactive attitudes are not a necessary or essential feature of blame on this view, Franklin seems clearly to embrace at least some form of functional reactivity. Because the aims of blame cannot be met without the negative reactive attitudes on this account of blame, this view is also likely to end up in blame curmudgeons’ crosshairs.

Before turning to the problem of blame itself, a few brief remarks on whether or not this examination of Hieronymi and Franklin’s accounts as test cases for reactivity suggests any broader conclusions about quality of will and functionalist accounts in general might be helpful. For my part, I expect that most functionalist accounts of blame are likely to be reactive accounts. That is because such accounts often appeal in some way to the kinds of dispositions needed to act as strong *motivating forces* for achieving

²³ Franklin’s view is, of course, more nuanced than this, in that he takes care to argue that proper valuing will often depend on a set of standards unique to the kind of thing being valued. When it comes to persons in particular, the negative reactive attitudes are crucial to upholding these standards. I will discuss these features of Franklin’s view in greater detail in Chapter 4.

the aims of blame. While conceptually it may be possible for other dispositions to serve this kind of role, I take it to be instructive that Smith, McGeer, and Franklin all coalesce around dispositions involving the negative reactive attitudes as the paradigm for this kind of motivating force. As Joshua Greene notes, “when Nature needs to get a behavioral job done, it does it with intuition and emotion wherever it can” (Greene, 2010: 367).

When it comes to quality of will accounts, my expectations are similar. Hieronymi’s account is not unique in taking the negative reactive attitudes to be not only closely correlated with evaluative judgments of ill will, but to perhaps have some tighter connection to them when it comes to our ability to mark their significance. And so it would not be surprising if the majority of quality of will accounts turn out to be reactive and subject to the problem of blame as well.

But what precisely is this problem? While I have characterized it broadly in terms of a kind of tension that arises between the value blame seems to have and the harm that it causes, I have not yet fully explicated the problem that blame curmudgeons and others inclined toward skepticism about blame are concerned with. Now that we have a clearer picture of the kind of blame that gives rise to worries about permissibility in the first place, we can now move on to the task of more clearly explicating the problem of blame itself.

1.4 The Problem of Blame

What I have here called *the problem of blame* concerns whether we should, all things considered, ever blame one another. Questions about normative adequacy are essentially *prescriptive* in nature. We want to know what we ought to do when it comes to attributions of blame. Are these attributions, their corresponding blaming practices, and the broader responsibility-related practices that depend on them ever justified and permissible to engage in?

As should now be evident, the many varieties of blame and the corresponding differences in the harm that can accompany blame complicate the task of assessing where the bar for establishing permissibility lies. For some accounts of blame, the bar seems relatively low. Returning to Scanlon’s conative account, while some challenge the descriptive accuracy of his view (namely whether it fully captures the kinds of blame that appear to be central to our actual blaming practices), the bar for the permissibility of this kind of blame seems low. Because Scanlon does not take any affective response or expressive feature to be necessary to blame, his

account is consistent with full-blown blaming practices that could involve little to no harm.²⁴ It is only when the variety of blame on offer accommodates the particularly unpleasant features of blame (as McGeer puts it, “warts and all”) that the stakes are sufficiently high to give rise to the problem of blame. And reactive blame is precisely the kind of blame that raises the standards for permissibility the highest. But what, precisely, do these standards amount to? Here I will attempt to explicate them in terms of two desiderata for *normative adequacy*, making use of an analogy with a related problem concerning permissibility and harm when it comes to punishment.²⁵

1.4.1 Blame versus Punishment

The first step to explicating the desiderata for a normatively adequate account of reactive blame is by noting an analogy with a similar and more familiar problem: *the problem of punishment*. Like reactive blame, punishment causes harm. The problem of punishment arises when we try to understand how and why it could be permissible to intentionally inflict the harms of punishment on wrongdoers, despite our standing pro tanto reasons not to cause harm to others.²⁶ Two possible solutions to the problem of punishment – deterrence and retributivism – can each serve to elucidate something about the desiderata for a normatively adequate account of reactive blame.

On one hand, deterrence theorists sometimes argue that it is permissible to intentionally cause harm via punishment because the reasons we have not to harm can be *outweighed* by some overwhelming good. In particular, the harms of punishment might reduce harm on a larger scale. On the other hand, retributivists sometimes argue that it is permissible to intentionally cause harm via punishment because the reasons we have not to harm are *defeated* in some cases, namely when the relevant harm entailed by punishment is deserved. When it comes to the permissibility of

²⁴ One might even think that it would be downright puzzling to demand arguments for establishing the permissibility of withdrawing from someone who has clearly manifested ill will toward you.

²⁵ I credit and thank Vargas (2013) for this helpful way of characterizing the relevant standard.

²⁶ Here, I assume that there are at least some moral reasons, that among them is a standing pro tanto reason not to cause harm, and that this reason often entails that we ought to refrain from certain kinds of harmful behaviors. While I wish to avoid any commitments regarding the nature of our moral reasons, I take it as a given that anyone who is at least agnostic about the permissibility of blame will be happy to accept these assumptions. Those who wish to deny the existence of moral reasons of any kind will take any success theories of moral responsibility or blame as a nonstarter, and thus will fall outside of the bounds of my target audience in this book.

punishment, we might therefore characterize a possible desideratum for a normatively adequate view of punishment in the following way:

Normative Adequacy_(punishment): we are justified in punishing if either (a) our reasons not to punish are outweighed because punishment entails some overwhelming good, or (b) our reasons not to punish are defeated because in at least some cases the harm entailed by punishment is deserved.

Might we say the same thing about the desiderata for a normatively adequate view of reactive blame? Here the corresponding standard would be that we are justified in blaming if either (a) our reasons not to blame are outweighed because reactive blame entails some overwhelming good; or (b) our reasons not to reactively blame are defeated because in at least some cases blame is deserved in the basic sense.

While this first pass is instructive, I do not think that it is ultimately correct. This is due to an important difference in the relation between (a) and (b) for punishment versus blame. In the case of punishment, deterrent and retributive strategies for establishing the permissibility of punishment are usually pursued independent of one another.²⁷ The desideratum for a normatively adequate account of punishment sketched earlier is in turn a disjunctive one. Punishment is permissible if *either* the deterrence theorist or retributivist gets things right. But this does not seem to be the case for reactive blame. Rather, when it comes to blame there is an underlying assumption that the value of blame could outweigh our reasons not to harm only if blame is *also* deserved in the basic sense, and that basic desert of blame could defeat our reasons not to harm only if the value of blame *also* outweighs our reasons not to harm.

To see this more clearly, consider first the way that whether or not the value of reactive blame outweighs our reasons not to harm is dependent on basic desert. Recall that, for example, on Franklin's functional account reactive blame is a necessary condition for valuing what we ought to value. If this is correct, then there is at least a *prima facie* case for thinking that the value of even reactive blame might outweigh the disvalue of its unpleasant and harmful features. When it comes to ranking the value of our various moral practices, properly valuing the objects of moral value themselves looks like the kind of thing that plausibly appears at the top of the list. But, even if this is correct then the following challenge is also plausible: the value of this practice (properly valuing the objects of moral

²⁷ This is, of course, an oversimplification in that one might also offer a *hybrid* account of punishment that incorporates both consequentialist and non-consequentialist considerations. See, for example, Hart (1968) and Scheid (1997).

value) will only be relevant to the *all things considered* normative adequacy of reactive blame insofar as reactive blame is also *deserved*.

If we want to know whether or not it is ever permissible to reactively blame, it is not enough to say that reactively blaming is good for us.²⁸ This is perhaps most obvious when considering the perspective of the unfortunate *targets* of reactive blame. Anyone who has ever been on the receiving end of misplaced resentment will agree that attempts to justify blame via appeal to the overall value of blame alone will be of little comfort. The harms of reactive blame raise the stakes. Reasons for thinking that the harms of reactive blame are outweighed by its value are not enough if we lack any further reason to think that those harms are also deserved. We must also have good reason for thinking that our standing reasons not to harm one another are *defeated* when it comes to reactive blame. What the intuitive impermissibility of misplaced reactive blame highlights is that appeals to value alone are insufficient to meet the relevant standard of normative adequacy without an accompanying defeater.²⁹

On the other hand, might appeals to basic desert alone be enough to do the relevant justificatory work? Mere appeal to desert also appears insufficient to do the relevant justificatory work. Assume for the sake of argument that we do have a defeater to the harms of blame – some people are in fact deserving of blame for some of their actions in the basic sense. Those skeptical of the permissibility of blame might still reasonably push back against the claim that reactive blame is normatively adequate. Surely there are things we might do that, while deserved, would not be all-things-considered permissible. Hitler, for example, might have reasonably deserved to live out the remainder of this natural life in a torture chamber. Had he survived WWII, this fact about desert in itself would not be sufficient to justify the claim that it would be permissible for anyone to actually lock him up in one and throw away the key.³⁰

²⁸ For articulation of a similar problem for revisionist views about free will, *the normativity anchoring problem*, see McCormick (2013). See also criticism of McKenna's account of the warrant of blame, his substantive desert thesis, in Chapter 2.

²⁹ While I take this point to be relatively uncontroversial for possible dissent, see Ciorria (2019). Ciorria argues that advancing intersectional feminist aims is sufficient to render blame apt, even if it is not deserved.

³⁰ Examples abound here, pick your favorite serial killer. There are also two possible readings of this claim. On the first reading, the claim is that it would not be permissible to lock him up. On the second reading, the claim is that even if this punishment is morally appropriate, no actual person would have standing to enact it. The point that desert does not on its own entail permissibility goes through on either reading. Thanks to Neal Tognazzini for pointing out this ambiguity.

The same can be said of reactive blame. Perhaps it turns out that the negative reactive attitudes are far more harmful to creatures like us psychologically speaking than we currently recognize. If we came to discover this fact, or that attributions of blame violate some other weightier moral considerations, then skepticism about the normative adequacy of this kind of blame would be reasonable despite the further fact that it is deserved in the basic sense. While the impermissibility of misplaced reactive blame suggests that basic desert is necessary for a normatively adequate account, further considerations like these suggest that it is not sufficient.

1.4.2 *Two Desiderata for Normatively Adequate Reactive Blame*

Defending the permissibility of reactive blame therefore requires meeting two desiderata for normative adequacy. One must defend both the claim that reactive blame is sometimes deserved in the basic sense, *and* the claim that the value of this kind of blame outweighs its corresponding harms. In order to meet blame skeptics on their own terms, a maximally persuasive argument for the permissibility of reactive blame should meet the following two desiderata for normative adequacy:

Normative Adequacy/value (NAV): we are justified in blaming only if our reasons not to blame are *outweighed* because blame entails some overwhelming good.

Normative Adequacy/desert (NAD): we are justified in blaming only if our reasons not to blame are *defeated* because in at least some cases blame is deserved in the basic sense.

I take each of these desiderata to be independent necessary conditions for establishing the permissibility of blame. Taken together, they are also jointly sufficient. Despite being unpleasant and even harmful, if reactive blame entails some overwhelming good for creatures like us *and* it is also deserved in the basic sense, then it is unclear what more skeptics about its permissibility might reasonably require.

Do we have any reason to think agents can be (and sometimes are) deserving of reactive blame in the basic sense, and that this kind of harmful blame is valuable? My goal in the next several chapters will be to systematically argue for affirmative answers to both of these questions.

1.5 Conclusion

Here I hope to have laid the groundwork for a defense of the permissibility of reactive blame. In light of the discussion in Section 1.1, we can proceed

with a clearer sense of what reactive blame is, and why skeptical concerns target this variety of blame in particular. Section 1.2 serves to further distinguish between the kind of blame that gives rise to the problem of blame and those that do not by canvassing some of the most prominent accounts of explicitly nonreactive blame. Section 1.3 offers a demonstration of how we might deal with borderline cases, those that do not clearly lie on either side of the fault line of reactivity. Finally, in Section 1.4, I explicate the problem of blame itself. What standards of normative adequacy must be met in order to meet blame curmudgeons on their own terms? Here I argue that a plausible account of normatively adequate reactive blame must meet both a value-based desideratum (NAV), and a desert-based desideratum (NAD). I now turn to the prospects for an account of reactive blame to meet the latter.