Civilizing Blame

Victoria McGeer

Introduction

The phenomenon of blame, understood most broadly as a distinctive class of negative reactions prompted by wrongdoing and directed toward wrongdoers, has lately received a great deal of attention from moral philosophers. Perhaps this is not surprising. Certainly blame, by anyone's reckoning, is a pervasive feature of our moral experience. Rightly or wrongly, happily or unhappily, human beings are the kind of creatures that go in for blaming one another for perceived wrongdoing, where wrongdoing involves not just doing a bad thing but doing something *wrong*—that is, doing something that the perpetrator knows, or ought to know, is a bad thing without any justifying or exculpating excuse. Of course, the wrong thing in question need not be a moral offense. I may blame you for failing to pay a bill on time, or weakly accepting an invitation I know you have good reason to turn down, or forgetting to take the dog out for its morning walk. Yet as Bernard Williams points out, even in such cases the wrongdoing (act or omission) has "some kind of ethical dimension to it"; that is to say, "for blame to be appropriate, there must be some generally reprehensible characteristic involved in the explanation: the agent must have been careless, or lazy, or self-serving, or something of the sort" (1995a, p. 40). Such cases shade by degrees into those more centrally associated with what might be called 'moral blame,' where blamers perceive the perpetrator's wrongdoing as demonstrating a morally significant measure of disregard or disrespect toward themselves or toward others whom the perpetrator wrongs.

It is blame in this moral sense that has attracted the attention of an impressive array of moral philosophers, including (among others) Nomy Arpaly (2003,

Previous versions of this paper have been presented at a number of colloquia and workshops, and I am very grateful for all the challenging questions and comments I have received on these occasions. I owe a particular debt of gratitude to those who have provided me with more extensive comments, including Pamela Hieronymi, Daniel Kelly, Owen King, Nicola Lacey, Philip Pettit, Jesse Prinz, Tamar Schapiro, George Sher, Seana Shiffrin, Michael Smith, Manuel Vargas and, last but not least, the editors of this volume, Justin Coates and Neal Tognazzini.

2006), Pamela Hieronymi (2004, 2007, 2008), Gideon Rosen (2006), T. M. Scanlon (2008), George Sher (2006), Angela Smith (2007, 2008a), Jay Wallace (1994, 2011), Gary Watson (2004), and Susan Wolf (2011). One specific reason for this focus derives from a basic tension. On the one hand, because moral community depends for its existence on relations of mutual respect and goodwill, upholding the norms that constitute such relations—and so reacting negatively toward transgressors—seems part and parcel of what we owe to one another. On the other hand, the negative reactions characteristic of blame often have an angry, punitive edge that many theorists find morally distasteful, not least because of the range of painful, even destructive behaviors it so often motivates. In other words, there seems to be an unacceptable normative tension between recognizing the importance of blame as a negative response to wrongdoing and acknowledging various features or aspects of it that appear to be part and parcel of the way blame is expressed or enacted by real people in their everyday interactions. The consensus among moral philosophers is that this tension must be resolved if blame is to be condoned as playing an appropriate normative role in responding to wrongdoing.

The broader philosophical project that drives much of this work on blame is therefore one of showing how it can be a normatively acceptable, even valuable response to wrongdoing. I call this the project of 'civilizing blame' and agree with moral philosophers that it is an important one to pursue, but not, I claim, at the expense of taking the bile out of blame. In my view, this contravenes a constraint that any satisfactory account of blame must meet—that it should direct us to a psychologically plausible phenomenon, however unsavory, that answers to our ordinary conception of blame: it satisfies the ordinary connotations of the term, including connotations of resentment and anger and the desire for payback. The challenge is to accept this constraint while showing that blame may nevertheless be a normatively fitting response to wrongdoing.

I argue for an account in this paper that can, I believe, rise to this challenge. It is psychologically realistic in taking blame to be a phenomenon that displays the negative profile (warts and all) that is typically associated with blame. Yet it satisfies the normative constraint by explaining how this affective complex, unsavory as it undoubtedly is, can still do valuable normative work so long as it is constrained by social and institutional practices that support its more constructive features. I present and defend that view in the second and third sections of this paper. However, before coming to that argument, I lay out the rather different approach to civilizing blame that is adopted by many of the moral philosophers I named earlier. On their version of the project, civilizing blame amounts to sanitizing it—in effect, to purifying blame of what they take to be features they find normatively problematic; thus, by my lights, it is a psychologically unrealistic theory. Defenders of the sanitizing approach think, further, that this psychologically purified version of blame is normatively satisfying so far as certain features are eliminated. I say they are wrong on this count too. In rushing to reform the psychology of blame, they actually misrepresent what

is normatively problematic about those features and so end up with an account of blame that fails to meet either psychological or normative desiderata.

The paper is in four sections. In the first, I look at the sanitizing approach, situating the question addressed, setting out the concerns that motivate it, and raising my critical worry that this approach is psychologically unrealistic—and, for that reason alone, must be normatively unsatisfying. In the second section, I argue in support of my contention that it is psychologically unrealistic to identify blame with the purified phenomenon that sanitizing theorists target. In the third section, I show that the warts-and-all psychological phenomenon in which blame consists has, despite its problematic features, a rich normative role unmatched by the sanitizing counterpart. And then in a short conclusion, I advertise a further benefit of my approach, showing that the argument points to a plausible project in the institutionalization of blame; broadly, it supports something like the restorative approach to issues of criminal justice. This conclusion drives home the idea that any sensible prospect for civilizing blame lies not in transforming (*per impossibile*) the psychology of blame but rather in supporting better practices of blame in both formal and informal contexts.

1. The Sanitizers' Approach to Civilizing Blame

There are two distinct projects that can be discerned in the philosophical work on blame—projects that are related but can nevertheless be distinguished. The first of these is a project of “metaphysically justifying” blame: showing that there is no metaphysical bar to satisfying an essential condition of warranted blame—namely, that individuals may be genuinely culpable for the wrong that they do. Though important, this project is not the focus of this paper, and I hereby put it aside. My concern is with a project that, by contrast, simply presupposes individuals can be culpable in a way that is metaphysically unproblematic.¹ This project—civilizing blame—is focused on a different kind of question—namely, is there something in the nature of blame *itself* that makes it a morally objectionable response to (culpable) wrongdoing? Philosophers who want to civilize blame aim to show how and why this is not the case.

As I said at the outset, I am sympathetic with the fundamental goal of this project but object to a standard way of pursuing it. The standard strategy aims to develop or defend a plausible account of blame that preserves its normative powers while purging its unsavory characteristics. It is a strategy of “sanitizing blame.” And though I concede it has some initial appeal, I claim it will ultimately prove forlorn.

To understand its initial appeal, consider two questions that philosophers must address in pursuing their work on blame: (1) a conceptual question—What exactly

¹ Some might say these projects bear on one another in important ways, so should be tackled together. I don't dispute their interconnection (discussed further in the current section, see also McGeer 2011) but still claim we gain some analytic clarity by pulling these enterprises apart.

is blame?; and (2) a normative question—Is this phenomenon, so understood, morally defensible? The challenge, as they construe it, is this: if giving an intuitive or folk-psychological answer to the conceptual question seems to push toward giving a negative answer to the normative question, then defending a positive answer to the normative question will require giving a (somewhat) revisionist answer to the conceptual one. Specific theorists differ as to how revisionist an account of blame must be in order to pass normative muster, but none can go too far without vitiating the whole enterprise. After all, if the proposed account of ‘blame’ fails to connect in substantial ways with our ordinary understanding of what is involved in that phenomenon, then this is tantamount to agreeing with those who want to claim that blame as normally experienced has no positive normative standing; hence, it should be banished (to whatever extent is possible) from our moral-psychological repertoire (see, e.g., Smart 1961; Pereboom 2001).² Thus philosophers whose aim is to civilize blame by way of sanitizing it face two distinct challenges. First, they need to argue that their preferred account of blame is not *too* revisionist: it captures the central or, indeed, “essential” features of blame as a moral-psychological phenomenon, even if that means cleaning up the ordinary conception to some degree; that is to say, their preferred account passes psychological muster. Second, they need to argue that blame understood in their preferred sense is indeed “fairly” or “properly” directed at a wrongdoer in light of the wrong he or she has done; that is to say, their account passes normative muster as well.

Jay Wallace gives a nice illustration of this balancing act in his recent work on blame.³ In his estimation, the philosophical project (civilizing blame) should take the form of “find[ing] space between two extremes for a defensible form of blame to occupy” (2011, p. 348). At one extreme is a weakly evaluative view that sees blame simply as a matter of sizing up a wrongdoer's moral shortcomings as displayed in his or her acts and attitudes. Blame, on this view, is nothing more than a form of judgment, coolly registering facts about the wrongdoer independent of any response the blamer might make to the facts so registered (for defenders of this view, see, e.g., Smart 1961; Arpaly 2003).⁴ But despite some dissenting opinions, most philosophers would agree with Wallace that such an account misses blame's “characteristic (if elusive) quality of

² It should be noted that, for many philosophers—Smart and Pereboom among them—the primary reason for denying that blame has any positive normative value stems from the metaphysical conviction that no person is genuinely blameworthy for their actions; hence, to blame them for what they do (in any deep moral sense) is manifestly unwarranted.

³ This paper (Wallace 2011) is largely a critical response to Scanlon's view on blame, which, in Wallace's estimation, errs on the side of being too revisionist. See also Wolf (2011). I am largely in agreement with this aspect of Wallace's insightful paper (including his emphasis on the affective dimension of blame). Nevertheless I regard Wallace and Scanlon as pursuing (importantly different) variations on a fundamentally similar theme.

⁴ Arpaly herself now acknowledges that this purely cognitive (or doxastic) account of blame is not fully adequate to everyday phenomenon. See Arpaly (2006).

opprobrium" (2011, p. 348). In other words, they hold that judging someone blameworthy and actually blaming them are important to distinguish on both psychological and normative grounds. To judge a person blameworthy is certainly to find fault with them, maybe even moral fault. But to make such a judgment requires only that the judger assess that person's acts and attitudes in a way that is sensitive to the requirements of norm-governed behavior; it does not imply having any reaction to *them* as a blameworthy person. So, for instance, a therapist or parent may judge her patient or child to be blameworthy in a purely fault-determining sense, without thinking it right to blame him. To blame someone, by contrast, is to have in addition some kind of negative reaction to them as a blameworthy person—a feature that many would say makes blame a normatively more problematic phenomenon, although much will depend on what blame is supposed to add to the phenomenon of merely judging someone to be blameworthy.

This brings us to the other extreme: an essentially punitive or retributive view that equates blame with punishment, or at least pictures blame as a hot motivational prompt for a variety of sanctioning behaviors. Again, in many philosophers' estimation, this analysis seems to miss its psychological target. As Wallace says, "[W]e can blame someone without undertaking to punish them or to impose unpleasant sanctions of some kind" (2011, p. 348). But even should this view have some psychological traction, it is certainly not one that moral philosophers would happily endorse. As Wallace representatively insists, an essentially punitive account of blame "deprives . . . [it] of its legitimacy as a form of interpersonal behaviour" (p. 348). He adds, "[M]ost of us would feel some discomfort about acting with the aim of punishing or imposing sanctions on individuals for their moral transgressions. This is not, one wants to say, an appropriate way for autonomous individuals to interact with one another" (p. 348).

Now I grant that neither of these extremes does justice to the rich phenomenon of blame. But despite the ingenuity of various philosophical accounts, I doubt we can come to a satisfactory compromise between psychological realism and normative acceptability by trying to occupy an elusive middle ground in the analysis of blame, preserving its "quality of opprobrium" while stripping away its unsavory features, identified in particular with the angry punitive face of blame. In my estimation, a realistic assessment of the psychology of blame suggests that it is a robust human phenomenon with many unattractive features—for instance, often entraining a desire to hurt or get back at the wrongdoer. Furthermore though the significance of this fact is disputed (see, e.g., Scanlon 2008; Sher 2006), blame is typically emotionally toned, manifested in a variety of expressive behaviors that will often be experienced as punitive by those to whom it's directed. We are, after all, emotionally reactive creatures, feeling hurt and even attacked by the angry or indignant emotions of others, even when those emotions are in some sense merited or

deserved and notwithstanding how blamers come to reveal such emotions to us (running the gambit, as it may be, from cold withdrawal or the silent treatment all the way through to verbal recriminations or physical threat; indeed sometimes a nasty or condemnatory look can be enough to make a wrongdoer suffer).⁵ The question is how to take all this on board without giving up on the overarching project of civilizing blame.

The remainder of this paper addresses this problem. In section 2 I set out a hypothetical, naturalistic account of the psychology of blame, arguing that emotion is intimately connected with blame and cannot be analyzed away. In section 3 I address two distinct worries that I see as implicit in the concerns of sanitizers about such an approach and defend the view that we can be naturalistically realistic about blame and yet see an important normative role that it plays. In section 4 I look briefly at institutional developments in criminal justice and related areas that would build constructively on this naturalistic understanding of blame.

2. Excavating the Complex Psychology of Blame

2.1 IS BLAME EMOTIONALLY TONED? WHAT A NATURALISTIC APPROACH CONTRIBUTES TO CONCEPTUAL ANALYSIS

Blamers characteristically experience a range of negative emotions toward those whom they blame: anger, indignation, resentment, outrage, irritation, frustration, hurt feelings, distress. To what extent are these negative emotions part and parcel of blame? There are those who favor a cooler account of blame, insisting that, while such emotions are often associated with blame, they are not essential to it (Scanlon, Sher). Others argue that blame would not be *blame* (at least as we ordinarily understand it) absent such emotional reactivity, especially absent the core emotions of anger, indignation, and resentment (Rosen, Wallace, Wolf). How is this dispute to be resolved?

Here is one argument that proponents of cool blame offer in favor of their position: Despite the fact that certain emotions are often, even characteristically,

⁵ This argues for taking a more encompassing view of sanctions according to which they are many and various, distinguishable in terms of a number of cross-cutting dimensions—for example, formal versus informal, substantive versus symbolic, material versus psychological. And though adults may judge it inappropriate to sanction one another in certain ways (they cannot tell one another to go stand in a corner), this still leaves a wide array of possibilities open to them. Indeed it is a sign of maturity that adults are susceptible to a variety of more subtle forms of sanction (the angry look, the cold withdrawal of friendly feeling). Hence though philosophers may be tempted to take a narrower view of 'sanction' (say, one that is material or substantive in certain key respects), I think that is psychologically unrealistic, and not just from the perspective of the one who is sanctioned but also from the perspective of the sanctioner (see also Sripada 2005). As P. F. Strawson (1962) points out, we are too invested in the goodwill of one another not to understand the chilling force of its withdrawal, and can use such withdrawal to powerful sanctioning effect (i.e., as a form of social regulation).

associated with blame, this is not invariably the case. Blamers may feel atypical emotions—or no emotions at all. Yet even in cases of “dispassionate” blame, where the blamer has no affective reaction to the wrongdoer as such, it seems intuitively obvious (at least to some) that the blamer fully blames the wrongdoer for something she has done. For instance, it seems intuitively obvious (at least to some) that we might fully blame a distant third party for some egregious act without any flutter from our amygdalas, as when we read about some crime in the morning newspaper. But, the argument continues, this implies that negative affect has no critical role to play in a philosophical account of blame. After all, philosophical analyses attempt to delineate the essential features of such phenomena. And since negative affect is not present in every legitimate instance of blame, it could not be an essential feature. Hence, philosophers should put it to one side in their attempts to specify what precisely blame involves.

In my view this argument is misguided, encouraging accounts of blame that are psychologically distorted. However, my resistance to it does not depend on contesting the claim that there may be genuine instances of dispassionate blame or instances of blame wherein the blamer has an atypical affective reaction to the person blamed (intuitions clearly differ on this point, for reasons I think will become clear in a moment). My concern is rather to challenge the guiding assumption that a proper analysis of blame should focus on identifying so-called essential features (features present in every legitimate instance of blame), discounting those that are “merely characteristic.” While this is a standard move in analytic philosophy, it can be overdone, showing insensitivity to the fact that characteristic features can be of two very different types. They can be features whose contingent association with the phenomenon in question is of no criterial significance. Call such features, ‘kind-associated features.’ Or they can be features that have criterial significance in the following sense: they account for our interest in identifying a kind as such, even though things belonging to the kind do not invariably manifest the feature in question. Call such features ‘canonical features’ (Jackson and Pettit 1995).

Consider how we come to identify the kind ‘heart.’ 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. Or consider how we come to identify the color kind ‘red.’ Things that are red in color are canonically the kind of things that produce a certain visual sensation in visually unimpaired human beings under normal lighting conditions. Something gets to be called ‘red’ only so far as it is a thing of *that* (visual) kind. But, of course, this doesn’t mean that all red things invariably produce the requisite visual sensation. Viewing conditions may be atypical in various exogenous or endogenous ways. And here is one final example, perhaps more relevant to the phenomenon in question. Happiness is canonically the kind of psychological state that

is affectively toned in a particular way. A person is only characterized as ‘happy’ so far as he or she is in a state of *that* (psychological) kind. But, at least if great writers are to be believed, there are occasions when individuals can be happy (and even know they are happy) without the canonical affective experiences. Jane Austen gives us a nice example of this at the dénouement of *Pride and Prejudice*, when, against all odds (including her own initial and well-articulated antipathy), Elizabeth Bennett has finally accepted Mr. Darcy’s hand in marriage:

The evening passed quietly, unmarked by anything extraordinary. The acknowledged lovers talked and laughed, the unacknowledged were silent. Darcy was not of a disposition in which happiness overflows in mirth; and Elizabeth, agitated and confused, rather knew that she was happy than felt herself to be so; for, besides the immediate embarrassment, there were other evils before her. She anticipated what would be felt in the family when her situation became known; she was aware that no one liked him but Jane; and even feared that with the others it was a dislike which not all his fortune and consequence might do away.

In keeping with these examples, I suggest that blame constitutes a moral-psychological kind, a canonical feature of which is its emotional tone. One of the attractions of this view is that it’s commensurate with a more general functionalist approach in the philosophy of mind. According to functionalism, folk-psychological concepts like ‘blame’ are best understood as designating states or attitudes that typically play a certain causal role in the cognitive economy of (typical) human beings (Lewis 1980; Armstrong 1980). In the case of blame, it is a state that is apt for being caused by perceived wrongdoing and apt for producing certain behavioral effects. (I will say more about these behavioral effects below.) Now on a functionalist approach, the fact that blamers characteristically experience a range of core emotions that color their blaming acts and attitudes (anger, indignation, resentment, outrage) strongly suggests two things. First, the state that typically plays the causal role of blame in human beings is an affective state, though perhaps not invariably so; there may be various deviating background conditions in place, as in the happiness example given earlier. Second, the emotional character of the state is not incidental to the causal role it typically plays; on the contrary, emotions are motivational drivers in human beings, reliably and involuntarily priming certain action tendencies. Moreover, because of this, or so I shall argue, their presence transforms the meaning of the actions thereby produced, giving those actions a particular normative cast.⁶

The approach I take in this paper is thus avowedly functionalist—and more generally naturalistic. I cite a number of advantages in the following pages. But

⁶ Cf. Wallace (2011) for a compelling conceptual defense of this point.

one at least is this: it helps tip the balance on a hotly disputed issue as to whether emotional reactivity is a constitutive feature of blame. By making sense of the fact that exceptional cases are beside the point, it lends credibility to the conceptual point that blame would not be blame absent the characteristic presence (and effect) of certain core emotions: anger, indignation, and resentment. Of course, as we know from everyday experience, the intensity of these emotions may vary widely, running the gambit from mild irritation all the way through to burning outrage. And the duration of these emotional episodes may vary as well, from relatively fleeting to unhappily persistent. Such variability surely depends on a number of factors: the type of wrongdoing; the frequency of the wrongdoing (does it represent an established pattern?); the wrongdoer's own attitudes and subsequent behavior; the blamer's own position vis-à-vis the wrongdoer; and so on. Yet even amid such variability, what I am calling the emotional tone of blame—the anger, resentment, indignation—remains relatively constant. So this naturalistic approach suggests a more fundamental question: Why is blame emotionally toned in this way? That is, why is the negative emotion canonically associated with blame some species of anger rather than, say, sadness (*pace* Scanlon)? This in turn leads to a more general question: Why is blame emotionally toned at all? What function does this serve?

These questions point up the need for thinking about blame in the wider context of understanding the role of emotions in human cognition and behavior. While this is a broad topic, philosophical work on the emotions has become increasingly interdisciplinary in recent years, deriving considerable insights from evolutionary biology, cultural anthropology, cognitive neuroscience, and experimental economics (see, e.g., Griffiths 1997; DeLancey 2001; Prinz 2004). Leaning on such work, I turn now to developing and defending a hypothesis that accounts for why anger is a characteristic feature of blame and, more broadly, why blame should have the kind of complex profile we see in everyday life—a profile that encompasses both problematic and positive normative characteristics. This account involves dipping into our evolutionary heritage, understanding that such reactions can be deeply embedded in human nature but still have a psychological complexity that is characteristic of a brain that has evolved in part by adapting old functions to new uses.⁷ While the evolutionary account I offer is provisional, allowing for considerable refinement and revision in its detail, it serves to make plausible my basic thesis that blame has certain elemental or primitive features on which later evolving cognitive sophistications have been built. More to the point, it serves to underscore how these features are inextricably intermingled, making sense of the everyday phenomenology of blame, while ultimately putting paid—as I shall argue—to the sanitizers' strategy for civilizing blame.

⁷ For a similar kind of approach aimed at evaluating the role of disgust in moral judgment, see Kelly (2011).

2.2 BLAME AS AN EVOLVED PHENOMENON: ELEMENTAL FOUNDATIONS

I begin with some preliminary remarks on the nature of anger. There are many theorists who contend that anger is a "primary" or "basic" emotion (along with fear, happiness, disgust, sadness).⁸ Basic emotions are distinctive so far as they appear to be universally shared among human beings, with cross-cultural patterns of eliciting conditions, physiological manifestations, facial expressions, and behavioral responses (Ekman 1972, 1992). This suggests that they are, at least in part, innately programmed responses to stereotypical situations in the world of relevance to individual survival and/or well-being. Such responses are rapid and involuntary, recruiting areas of the midbrain in the absence of, or at least prior to, any higher cortical activity; hence they are not cognitively mediated in a reflective or judgmental sense. Even so, they involve a *backward-looking element of appraisal*: they are triggered by various events only so far as those events are coded (i.e., perceived) as situations of the relevant emotion-inducing type. And they involve a *forward-looking action-priming element*, disposing agents to behave in ways that, in situations of the relevant type, had an overall tendency to enhance individual fitness in ancestral populations. For instance, thanks to our evolutionary heritage, we have an innate fear response triggered by looming objects. The response is rapid and involuntary, involving a noncognitive appraisal that looming objects are potentially dangerous and simultaneously priming a suite of defensive behavioral reactions (ducking, protecting vulnerable body parts, turning away, fleeing). This same general pattern applies to anger, only here the kind of situations that involuntarily provoke the emotion are coded (i.e., perceived) as something like "an offense to me or mine," and the response thereby provoked is a suite of aggressive behaviors aimed at defending whatever it is that's threatened by the offense.⁹

With these general observations in mind, I turn now to the question of why anger is involved in our blaming acts and attitudes. The first thing to note is that blaming anger is often triggered by something more specific than "an offense to me or mine"—namely, norm transgressions of one sort or another. Why should this be the case? An evolutionary approach would suggest that early hominids were subject to selective pressures that favored the policing of social norms. And indeed many theorists argue that large-scale human cooperation could emerge only with the development of social norms, supported

⁸ While different theorists often give a more extensive list of basic emotions, differing considerably on what should be included, they generally agree on the five listed here. For discussion, see Griffiths (1997); Prinz (2004).

⁹ For discussion, see Prinz (2004), especially chapter 3.

and defended by cognitive/affective mechanisms specifically recruited to this task (Nichols and Mallon 2006; Henrich and Henrich 2007; Boyd and Richerson 2005; Sripada and Stich 2007; Nichols 2002). Of course, various emotions might lend themselves to such recruitment; for instance, disgust is argued to be importantly involved in supporting social norms (Kelly 2011). Still, there can be little doubt that anger, and the suite of aggressive behaviors it entrains, should play a major role in this process, perhaps particularly with regard to supporting those norms involving fairness and harm (Sripada and Stich 2007).

In support of these evolutionary speculations, consider a convergent line of research from behavioral economics. A number of game-theoretic studies involving norms of social exchange demonstrate that punishment, and the threat of punishment, stabilizes cooperative interactions, especially when there is a tendency in the population toward "altruistic punishment" (the punishers are bystanders who have not been directly affected by the transgression) (Fehr and Gächter 2002; Fehr and Fischbacher 2004a, 2004b). Yet from the economist's perspective, the emergence of such a tendency is somewhat mysterious. After all, punishment is costly, not only using up valuable resources but also potentially incurring the costs of reprisal. So why would self-interested agents go in for such a thing, especially if it were not to benefit them directly? Robert Frank (1988) suggests the answer lies in our emotions (and specifically here in our blaming anger). On his account, thanks to the pressures of natural selection, human beings are not simply (or primarily) rational agents in a narrow calculative sense; we are also *passionate* agents—agents whose short-term behavior can serve broader rational goals when driven by forces that do not depend on more narrowly circumscribed reasoning processes. In Frank's terms, these forces—that is, emotions—not only prime us to act in ways that serve our broader rational interests; they operate as "commitment devices," making credible to others the overwhelming likelihood of our acting in such ways. Hence, the fact that human beings respond emotionally to things not only has an important action-priming function; it also has an important *signaling* function, a feature I will return to below.

The relatively simple story of blame that emerges from taking a broadly evolutionary perspective can thus be summarized as follows. At the most elemental level, blame is a quasi-autonomous, emotionally mediated response to others that is specifically prompted by, and targeted on, behavior that transgresses personally or socially valued norms. Despite its quasi-autonomous (nonreflective) character, this blaming response has both a backward-looking appraisal dimension and a forward-looking regulative dimension. The backward-looking appraisal dimension consists in coding—that is, perceiving—others' behavior as (offensively) transgressive; the forward-looking regulative dimension consists in an aggressive (punitive) response aimed at changing or inhibiting such behavior.

2.3 BEYOND ELEMENTAL BLAME

But is this simple story adequate to our everyday experience of blame? On the one hand, it certainly accounts for what appear to be the cruder features of blame, even in its contemporary guise. (Consider, for instance, the power and immediacy of "road rage"—the angry blame one feels at being cut off in traffic.) Yet, on the other hand, it seems to miss certain respects in which blame can function as a more reflective and nuanced response to wrongdoing, open—to some degree—to reflective criticism and control. For instance, we are generally well able to withhold or suppress our blame when we judge (for any of a number of reasons) that someone does not deserve such a response despite his or her transgressive behavior. How, then, to account for the more sophisticated ways in which our blaming emotions operate?

The answer, I believe, lies in our higher-order *mentalizing* capacities, the fact that we understand what others do in terms of a wide variety of mental states (beliefs, desires, intentions, emotions, and so on). Such capacities dramatically transform how human beings are able to experience blame, both as individuals who do the blaming and as individuals who are the targets of blame. They make sense of how a more primitive system of blaming anger could have been recalibrated, not simply in terms of the kinds of situations that would elicit it, but also in the kinds of responses to which it would naturally give rise.¹⁰ Nevertheless, recalibration is not wholesale replacement. Recalibration works by co-opting basic features that are already in place. Thus we might expect that such developments in the cognitive anatomy of blame would simply have built on its more elemental foundations. And this implies that blame remains, at root, an emotionally mediated response to another's behavior, structured in terms of a backward-looking appraisal dimension *and* a forward-looking regulative dimension.¹¹ How, then, has this two-dimensional focus of blame been transformed by our mentalizing capacities?

Consider, first, the backward-looking appraisal dimension. I take it to be undeniable that when we blame others it is principally for something they *do*: for their transgressive behavior. Yet as mentalizing creatures, it is also undeniable that we care deeply about the underlying attitudes that motivate and thereby explain what others do. We care about this not just because it tells us more about how to code some particular instance of behavior—to determine

¹⁰ For a suggestive account of how basic emotional systems become recalibrated by way of higher-order cognitive processes, see Prinz (2004).

¹¹ Fully acknowledging this forward-looking dimension of blame has the potential of removing a philosophical puzzle that seems to drive a lot of the work on blame: understanding the peculiar force or power of blame to do something more to the wrongdoer than give negative evaluation of his or her behavior. In my estimation, this force largely resides in the regulative dynamics of blame, which dynamics many philosophers are resistant to acknowledging because of what I shall argue are misplaced normative concerns.

if it was genuinely a transgressive act, as opposed to, say, an accident (e.g., he stepped on my foot deliberately, not just because he didn't see it). We also care about a person's underlying attitudes because we take these to be reliable indicators of *patterns* of behavior—that is, how that person is likely to act toward others, or again toward ourselves. It is thus our mentalizing capacities that explain P. F. Strawson's insightful observation that our blaming emotions—resentment, indignation, anger—are acutely sensitive to the *attitudes* that others manifest in their behavior, whether what they do displays a “good will, its absence, or its opposite” (1962, pp. 76–77). The focus on attitude explains, indeed, why Strawson calls our blaming emotions “*reactive attitudes*,” insisting that, even though we experience such attitudes when others do something transgressive, we generally endorse our reaction only when that behavior betrays some attitudinal deficiency on their part that has more general significance for their interpersonal activities (cf. Strawson's long discussion of excusing and exempting conditions, pp. 77–80).

Of course, such attitudinal deficiencies may be many and varied. In blaming, we might judge others to be lazy in their interpersonal concern, unthinking, indefensibly ignorant, neglectful, uncaring, disrespectful, mean, or downright malicious—or even some combination of these. Yet if my general thesis about the foundational structure of blame is right, the attitudinal focus in our experience of these blaming emotions should have both the backward-looking appraisal dimension *and* a forward-looking regulative dimension. That is to say, we should certainly care about what attitudes individuals manifest in their behavior. But if those attitudes are interpersonally deficient in some way, we should care equally about seeing a change in those attitudes, working to promote this change first and foremost through our blaming reaction. Indeed I think this is just what we find.

The two-way dimensionality of our attitudinal concern is made manifest, in my view, by thinking about what makes our blaming emotions go away, not in a purely psychological sense but rather in a normative sense—that is, what makes our blaming emotions feel “properly addressed” (so that if, psychologically speaking, we have any lingering remnants of anger or indignation, we judge these feelings to be misplaced or inappropriate).

On the backward-looking appraisal end of blame (to which moral philosophers have given most of their attention), our blaming emotions are appropriately modified if it turns out that the attitudinal deficiency we think we detect in someone's transgressive behavior is not in fact present. For instance, the person had a reasonable excuse for what she did: it was an accident, or she was compelled by circumstances beyond her control, or she was under great strain and did not mean to show any disregard or ill will, and so on. We feel mollified then, and our blaming emotions naturally tend to die away.

But equally, on the forward-looking regulative end of blame, these emotions are normatively answered—and naturally tend to die away—when

wrongdoers admit that their actions did display a bad attitude (that they were lazy in their concern, uncaring, disrespectful, mean, or even malicious); that such attitudes were, and are, inappropriate; that they are therefore responsible for the injury they caused; and, more important, that they take responsibility for not so harming again by changing or regulating the attitudes that led to the offensive behavior.

But how could wrongdoers convincingly show all this? In part, I claim, by their own reactive emotions of guilt, shame, and remorse and the activities such emotions tend to generate. These are themselves *reactive* emotions (or attitudes) because they react to the blaming emotions of others and, indeed, importantly invite a reactive response in turn—for instance, when all goes well, the blamer's mollification, understanding, and even forgiveness. Now I do not mean to suggest that this process is typically easy or straightforward and occurs in the blink of an eye. My point is rather to highlight a dynamic trajectory of unfolding events, orchestrated back and forth between blamer and wrongdoer, which we recognize as having a normatively appropriate, even desirable structure because of the potential power it has to develop the moral understanding of all parties in a normative dispute—most importantly, the offender's, but often as well the moral understanding of those who would call the offender to account. Importantly, this structure, which I have elsewhere called ‘reactive scaffolding’ (McGeer 2011), is enabled and enhanced because we are emotionally sensitive creatures, attuned to the way others are speaking to us in large part by way of their emotional responses. (I return to this point below.)

Understanding this feature of our psychology paves the way for restoring the credibility of the sanctioning face of blame; in particular, it gives us the necessary resources for responding explicitly to the kind of worries that may drive the sanitizers' quest to transform blame into something rather more civilized at the psychological level. In the following section, I explicitly identify and discuss two such concerns, aiming to show that they are fundamentally misguided, however well motivated they may be. My conclusion will be that somehow we must search for resources outside the head if we are ultimately to succeed at the task of civilizing blame.

3. Replies to the Sanitizers' Normative Concerns

The account of blame I offer in the previous section has two distinct goals. The first is to remain true to the complex psychology of blame as experienced in everyday phenomenology. The second is to give some account of that psychology within a broadly naturalistic framework. These two desiderata are met, I believe, by understanding how blame could be an evolved reaction to transgressive behavior (as it undoubtedly is, no matter what the correct details of such an account), with more nuanced reflective features (courtesy of our

mentalizing capacities) that build on a more primitive foundation. I leave any further defense of this account to another occasion.

My purpose now is to consider why the moral philosophers discussed in section 1 might find the nature of this bricolage somewhat disturbing, for, if I am right, the persisting elemental features of blame here identified are precisely what give rise to their deepest normative concerns. There are two concerns in particular that are likely to worry them. The first and most important of these underwrites their deep resistance to the sanctioning face of blame; I label it the 'anti-regulation concern.' According to this concern, a civilized form of blame cannot aim at the regulation of behavior without becoming an essentially conditioning enterprise in which the blamer fails to treat the one who is blamed with appropriate moral respect (i.e., fails to acknowledge his or her moral personhood appropriately). The second normative concern is less widely shared but tends to be articulated in addition to the anti-regulation concern, perhaps because of how blaming emotions so often motivate explicit sanctioning activities. I call it the 'anti-emotion concern.' According to this concern, while blame is often associated with angry, punitive emotions, these emotions are dangerous and surely not essential to a more civilized form of blame. After all, a civilized form of blame would simply consist in making a normatively appropriate response to wrongdoing—that is, a response that, however specified, is essentially nonsanctioning (in keeping with the anti-regulation concern). Why can't this be accomplished in an affectively neutral way? I try to address these two concerns in this section and use my responses to build up the naturalistic picture of blame introduced in the previous section.

3.1 THE ANTI-REGULATION CONCERN

The anti-regulation concern gives expression to the worry that there is something essentially disrespectful about blamers working to regulate a wrongdoer's behavior; hence a morally defensible account of blame cannot take such regulation to be an essential feature of blame. Why should this be the case? Here is a line of reasoning that may be driving the concern: To view blamers as working to regulate the behavior of wrongdoers is to see the blamers as engaging in an essentially disrespectful engineering enterprise, strategically undertaken to control the behavior of those who offend against them.¹² In other words, it involves treating the targets of blame not as moral agents in their own right but as objects of behavioral conditioning and control—in Strawson's

¹² This concern is explicitly voiced by Philip Pettit (2001). His solution to this problem is somewhat different from my own; it appeals to a notion of tacit consent to be so regulated by others. I go a somewhat different route, but our paths run parallel to the extent that we claim that discursive modes of affecting others' attitudes are essentially nonmanipulative; hence blame will be nonmanipulative to the extent it legitimately counts as a discursive mode of behavioral regulation. I come to this argument shortly.

terms, it is to take an "objective attitude" toward wrongdoers. But, so this reasoning continues, it is part of the very essence of our blaming activities that we treat wrongdoers as responsible moral agents in their own right. After all, if in appraising what someone does, we come to see them as out of their minds, or out of control in some way that demonstrates a compromised capacity for responsible agency, we think our blame—our blaming emotions—are misplaced. (These are the exempting conditions that Strawson emphasized.) So, as blamers, we cannot do any regulative (i.e., conditioning or engineering) work with our blame without undermining or betraying the very feature of our practice that makes such blame appropriate: treating wrongdoers as responsible agents in their own right, deserving the respect of moral address even if the content of that address is to express our anger or distress at their moral shortcomings.

I have two responses to this anti-regulation concern. The first relies on what I think is a fairly uncontroversial point and clears up an important ambiguity in the reasoning presented above. Some practices may have an underlying functional rationale such that we would surely give them up if they failed to fulfill that function. Nevertheless, fulfilling that function may not be what we explicitly aim at when we engage in the practice. Fulfilling that function may be an essential and, indeed, explicitly recognized by-product of the practice in question; yet it is not, perhaps cannot be what practitioners aim at without spoiling the functional efficacy of the practice in question.¹³ This familiar observation is what lies behind the hedonistic paradox: happiness is not something we can or do achieve by directly seeking it; yet we often engage in particular practices that aim at something else because we recognize that they make us happy, and we would give them up if they did not. So too it may be with the practice of blame. Blaming is something we do not with the explicit aim of regulating others' behavior. Nevertheless, in cases where we know blame can have no such effect, it loses its appeal as an appropriate or fitting response to wrongdoing (as when we say "It's no good blaming Jones; she's out of her mind"). This would make regulating others' behavior essentially a by-product of the practice of blaming, indeed one we explicitly recognize when we regard genuine insensitivity to the demands implicit in blame as a telling sign of the kind of problem (e.g., mental disability) that appropriately exempts others from our blaming attitudes and activities.

The "essential by-product" response shows that a regulative account of blame need not flout the normative prohibition against treating others in an essentially disrespectful way—that is, in a way that fails to respect their status as responsible agents. Strawson himself emphasizes this point in the final passage of "Freedom and Resentment":

¹³ The notion of some feature being essentially a by-product of a practice (i.e., not one that can be intentionally or explicitly aimed at) comes from Elster (1983).

It is far from wrong to emphasize the efficacy of all those practices which express or manifest our moral attitudes, in regulating behaviour in ways considered desirable; or to add that when certain of our beliefs about the efficacy of some of these practices turn out to be false, then we may have good reason for dropping or modifying those practices. What is wrong is to forget that these practices, and their reception, the reactions to them, really are expressions of our moral attitudes and not merely devices we calculatingly employ for regulative purposes. Our practices do not merely exploit our natures, they express them. Indeed the very understanding of the kind of efficacy these expressions of our attitudes have turns on our remembering this. (1962, p. 93)

Still, while this passage clearly endorses a view of blame according to which behavioral regulation can be an essential (and recognized) by-product of a normatively appropriate practice, I think it also suggests a stronger, more controversial view that I am inclined to take. So here is my second response to the anti-regulation concern. I claim there is nothing per se incompatible about consciously and explicitly aiming at changing (i.e., regulating) others' behavior and fully respecting them as persons—as responsible moral agents in their own right. What makes these things compatible or incompatible is the *means* one chooses to change or regulate the behavior in question.

Let me demonstrate this point first by way of considering an analogy. Suppose I believe that *p* is true and I want you to come to believe it too. (Of course, I may have a variety of motives for wanting this, but I don't think this potential variety affects the point I want to make.) Suppose, further, that there are a variety of means by which I could accomplish this goal: hypnosis, drugs, neural tinkering, or the good old-fashioned method of offering you what I take to be compelling evidence and/or argument in favor of *p*. Now even though I am consciously and explicitly aiming at getting you to believe that *p*, when I offer you evidence and argument to the best of my ability, no one would claim there is something essentially reprehensible in what I am doing, whereas if I adopted one of these other means, there certainly is. The difference is clear. In the argument case, I respect your own capacities as a reasoner to form (what to my mind are) true beliefs; I respect you as a reasoning agent. In these other cases, I treat your state of mind as something to be *engineered*, short-circuiting your own powers of reasoning to arrive at the truth of *p*. Note too that when I adopt such means, it does not matter whether my motives are evil or benign. Either way, there is something I am failing to respect about you *qua* believer—namely, your own capacity to track and evaluate the reasons for belief, a capacity that is central, as most would agree, to your autonomy and moral dignity as a person. Thus even were I concerned to act in your own best interest, I could not take the engineering approach to altering your beliefs without thereby compromising those interests. This is what makes the engineering approach morally reprehensible.

In sum, to be respectful of you *qua* believer is to be respectful of you *qua* reasoning agent. But in order to be respectful of you in this way, it does not matter that I explicitly aim at getting you to change your beliefs; what matters is that I choose a means whereby your own rational faculties are the proximate cause of the change in your beliefs. That is to say, I must offer you argument and/or evidence in favor of *p*. Of course, there may be ways in which I can fail to be normatively appropriate in offering you my arguments. (I use rhetorical tricks, I don't give you all the evidence at my command, and so on.) Nevertheless because *you* are the final arbiter in how you respond to my arguments (believing *p* or not), though I may mislead I do not directly and explicitly compromise your status as a reasoning agent. One significant and important test of this fact is that you not only have the power to withhold your belief, but you have the power to challenge my arguments and my evidence, thereby exposing me to the very same process and possibilities to which I expose you—specifically the possibility of changing *my* mind as to the truth of *p* in light of your argumentative response. Call this the 'dialogical test' on a normatively acceptable means for altering a person's beliefs.

Does this provide a good analogy for thinking about the case of blame? Specifically, can we use such reasoning to lay to rest the more direct version of the anti-regulation concern which holds that, in blaming others, we cannot explicitly and consciously aim to regulate their behavior in blaming them without disrespecting them as persons, as responsible moral agents in their own right? I think we can. To see this, recall first the mentalizing point: that what invokes our blaming attitudes is not transgressive behavior per se but the attitudinal deficiencies we take such behavior to display. Hence in blaming others, though we may aim to alter (or regulate) their behavior, it is by way of changing their attitudes—or, more precisely, by way of rectifying what seems to us a *deficiency* in those attitudes. Moreover, we may aim to change others' attitudes by way of communicating certain things to them through our blaming emotions, thereby stimulating their own capacity to track what is required of them as responsible agents. Specifically, our anger and distress communicates to others how we are interpreting what they have done, what it reflects about them, and that we don't find it acceptable. Further, by treating them as fitting or appropriate targets of our blame, we communicate to others that we regard them *as responsible agents*—hence that we have a right to expect better of them, both now and in the future, and that they will remain fitting targets of our blame absent reparative action or change on their part.

Now some philosophers may regard these blaming emotions not as a powerful communicative tool but rather as a brutally regulative stick. After all, ill will is not something that individuals enjoy having directed at them—indeed, as I emphasize in note 5, they can experience these attitudes as deeply punitive. This is simply the consequence of being (in Strawson's terms) reactively sensitive creatures, involuntarily disposed to react to the good or ill will of others,

particularly when that good or ill will purports to comment on the quality of one's own actions and attitudes. Still, this psychological fact, so far from making putative wrongdoers prey to the engineering control of those who blame them, draws them into an exchange about the precise meaning of what they have done. Moreover, because such exchanges have the power to make a blamer's blame feel properly addressed, this is—consciously or unconsciously—what blamers are asking, maybe even demanding wrongdoers to do: to explain and justify themselves and, failing that, to apologize or make other reparation and to rectify their actions and attitudes going forward.

This is just what makes the belief case an apt analogy. On the account I offer, the point and power of blame is to draw wrongdoers into a kind of exchange where they are perforce challenged to exercise their capacities as responsible agents, to reflect on what they have done, whether or not it is legitimate, and if it is not, to *take* responsibility for what they have done and for what they will do in the future. Thus despite the fact that such a process aims to change a person's actions and attitudes, it is not by means of the blamers' exercising engineering control. In the first place, this process aims to change wrongdoers' attitudes and actions by way of challenging them to better exercise their own capacities as a responsible agent: via the demand for explanation, justification, reflection, and reform. But in the second place, drawing putative wrongdoers into such an exchange, and thereby fully crediting them with the capacities of responsible agency, opens blamers themselves to being challenged on the legitimacy of their blame: it opens them to explaining and justifying their normative expectations and how they think those expectations have been unjustifiably disappointed by the targets of their blame. In other words, reactive exchanges of this sort are invariably dialogical: they invariably work to develop the normative understanding of the putative wrongdoer, the blamer, or both. Hence this process—which I have elsewhere called 'reactive scaffolding' (McGeer 2011)—passes the dialogical test on a normative acceptable means for attitudinal and behavioral regulation; indeed, it is because it passes this dialogical test that it is a particularly effective form of such regulation. To quote Strawson once again, "[O]ur practices do not merely exploit our natures, they express them. Indeed the very understanding of the kind of efficacy these expressions of our attitudes have turns on our remembering this" (1962, p. 93).

3.2 THE ANTI-EMOTION CONCERN

Finally, I return to the anti-emotion concern, which now may be put like this: Suppose the foregoing account is on the right track. Blame works to instigate a process of normative reflection on the appropriateness or inappropriateness of what someone has done, with the aim of shaping or improving the actions and attitudes of a genuine transgressor—a process that, by its nature, exposes blamers themselves to the potential normative review of their blaming actions

and attitudes. But now, if this is what accounts for the normative force of blame, why can't this process be carried forward, perhaps more effectively, without the coloring of a lot of (or indeed any) angry emotion?

There are two ways of interpreting this question. The first is a question about psychological reality: Is it possible for creatures like us to engage in such a process under the burden of perceived injury without feeling or expressing angry emotions, however mild these may be? My response to this is brief. Perhaps it is possible, at least for the saints among us. But however much we may suppose this is a preferable state of affairs, I doubt it is practically available to the common run of humanity. And this must pose a serious challenge to any philosophical theory that recommends it.

Still, for the sake of exploring the shape of our normative commitments, it may be worth posing this question in a second, more theoretical vein. Suppose we could dispense with the emotional concomitants of blame. Wouldn't this be a normatively preferable state of affairs? After all, we are all well-versed in the dangers of angry emotions. They can be highly destructive, leading to cruel and excessively punitive behavior. They can also be self-destructive, leading to unpleasant ruminations, loss of trust, loss of more general goodwill, loss of a capacity to appreciate the better things around us. They can get in the way of listening to what others have to say. They can generate downward spirals, where one person's anger begets what may seem to others justified anger and resentment in turn, refueling the anger originally felt and encouraging more anger in return. The list goes on. Hence even if the goal is unreachable, we should still work hard at tempering or repressing the angry concomitants of blame.

I concede all these dangers and do not want to downplay the difficulties of dealing with blame as an emotion-laden psychological phenomenon. Indeed, as I will emphasize in my next and concluding section, this is part of my motivation for writing this paper. Nevertheless, I claim that, dangerous as indignation, resentment, or outrage may be, the state of affairs in which such emotions are missing or absent from blame should not be recommended on normative grounds. In order to establish this point, I advert once again to the kind of creatures we are, for in my view, it is psychological reality that makes emotionless blame not just practically unavailable but also a misguided normative ideal. My claim turns on a premise I have defended in response to the anti-regulation concern: that, as a matter of psychological fact, emotions constitute for us a powerful form of communication. Here I up the ante by arguing that emotions constitute for us a *uniquely* powerful form of communication, the specific features of which militate against any normative recommendation that we should suppress them (to whatever extent possible) in contexts that make blame appropriate. I offer two considerations in support of this claim.

The first consideration focuses on how emotions make salient for us the communicative contents of what they convey. I have already stressed that, as reactively sensitive creatures, we are involuntarily disposed to react to—that is,

both detect and *care about*—the good or ill will that others manifest toward us. And this will be so, perhaps even more strongly so, when that good or ill will is itself a *reactive attitude*—that is, when it is directed toward us in virtue of how we have treated others (or more precisely, in virtue of how they judge that we have treated them). Such emotional reactions are attention-grabbers for us, providing a form of triangulation that draws our attention both to the person manifesting the attitude and to the behavior—whatever it was—that prompted the attitude. The attitude itself is usually self-intimating; we are very good at reading the emotional reactions of others, including—in the case of blame—the opprobrium it conveys. But that someone should have such a reaction draws our attention to *her*—perhaps in a way we had not given it before; it draws our attention to her not just as a particular person but as a creature that can be hurt by what we say and do and, as such, is deserving of our regard. Furthermore, while I have said others' blame directs our attention to our own behavior, it may not be obvious to us why our behavior should prompt such a reaction. Nevertheless, our own emotional reactivity makes this an issue we cannot easily ignore. Of course, we may resist their review of our behavior, getting angry or indignant in return, thereby conveying to the blamer that she has misjudged us or somehow reacted inappropriately. But even as we offer such resistance, we are willy-nilly drawn into an exchange with her about the meaning of our behavior, the nature of appropriate expectations, the shape of prevailing norms, and so on. And as I have repeatedly emphasized, we are drawn into this content-rich exchange precisely because of the quality of our own emotional reactivity to the emotional-laden communications of others.

The second consideration digs more deeply into why we should be so reactively sensitive to (i.e., particularly affected by) the emotional reactions of others, both from an evolutionary and a game-theoretic point of view. The suggestion is that emotions are what biologists call 'expensive signals': they are hard to generate under conditions that do not naturally prime a person to experience them; equally, they are hard to fake (Frank 1988). Blaming emotions thus have a special communicative (and evidential) power. Since they are hard to generate under conditions wherein blamers do not genuinely perceive an injury, they convey to transgressors both how seriously blamers regard the offense and their sincerity in pressing their normative demands. Furthermore, from the blamers' point of view, it is the emotional quality of the transgressors' reactions that conveys how seriously they are taking the blamers' complaint, as something to which they are sincerely paying attention (whether they accept that complaint by way of guilt, shame, or remorse or repudiate it by way of their own indignant anger). As Jay Wallace points out, emotions have the power to transform the meaning of our various communications; the suggestion here is that this is because emotions are not normally or straightforwardly under our voluntary control.

Taking these points together, we reach a conclusion that may not seem entirely satisfactory: although the anti-emotion concern is based on valid

normative considerations, these should not be trumping considerations. If we do not wear the costs of angry blame, then we cannot reap its considerable rewards. This is not a wholehearted endorsement of what blame has to offer by any means. But it is to insist that we cannot—and, more to the point, should not—deal with these normative concerns by recommending that we try (as I think, *per impossibile*) to take the emotional guts out of blame.

4. Civilizing Blame: Sketching a New Way Forward

I began this paper by identifying what seems to be an unacceptable normative tension between recognizing the importance of blame as a negative response to wrongdoing and acknowledging various features or aspects of it that appear to be part and parcel of the way blame is expressed or enacted by real people in their day-to-day interactions. The bad news of this paper is that this tension is irresolvable, at least at the psychological level. While there is some reason to regret the anger characteristically manifested in blame, and while we as individuals may try to repress or overcome its more punitive dimensions, we are fighting against a very deep feature of human psychology—a feature, moreover, that (in its more constructive guise) plays a critical role in regulating behavior by way of making salient the demands that shared norms place on our actions and attitudes. Hence, I conclude, the sanitizer's strategy for civilizing blame is not only psychologically forlorn; it is normatively misguided. If this project is not to be abandoned tout court, we need to find a new way forward.

The alternative I propose does not involve reshaping the psychology of blame, but it does involve thinking about how that psychology is given expression through our social institutions and practices of blame, some for better, some for worse. The guiding idea here is that, despite our liabilities, we are at least an engineering species, able to develop practices and institutions that serve our various ends in light of our shared human tendencies. Hence in pursuit of the normative goal of civilizing blame, it makes sense to ask which practices and institutions best serve the end of capitalizing on blame's more sophisticated regulative characteristics, while checking, or reining in, those characteristics that have an unsavory and potentially destructive punitive edge. I call this the strategy of *domesticating* blame. And while I think it may be pursued with some profit in a number of different domains (both public and private), the criminal justice system provides a particularly compelling site for this inquiry, since it is here that blame receives one of its most powerful and communally sanctioned forms of institutional expression. I therefore close with some very brief and sketchy remarks on what this strategy for civilizing blame might recommend, as well as caution against, in the field of criminal justice.

Recall that, in my view, what makes blame a normatively acceptable, even valuable response to wrongdoing is that it is no longer just a primitive punitive

response but one that builds onto this crude regulative structure a cognitively sophisticated *dialogical* dimension. Blame *responds* to something a wrongdoer has done and it *calls for* something from the wrongdoer in return: namely, that the wrongdoer appropriately address the blamer's sense of being wronged.¹⁴ Of course, how this response should go is something of an open question. Blame (in its most constructive guise) will be satisfied by a variety of responses, including a discussion of fault, the offering of excuses or exculpating explanations, a recognition of the blamer's right to better treatment, remorse, apology, restitution, commitment to reform, and so on. These things constitute a family of suitable responses to blame by taking account of who was wronged and how, while at the same time reinforcing our shared commitment to the norms that make for moral community. Hence blame (in its most constructive guise) aims to generate a normatively valuable dialectic, both from the point of view of repairing the variety of wrongs done by any transgression (to the victim, to the stability of shared norms, to a sense of security, and so on)¹⁵ and from the point of view of regulating behavior going forward. More important, I claim, what makes blame particularly apt for generating this dialectic is that it is the sort of phenomenon to which wrongdoers, as reactively sensitive creatures, are emotionally primed to respond. But could such features, which seem to require a more immediate form of encounter between blamer and wrongdoer, be given any true institutional support in the area of criminal justice? And even if they could, is this indeed a normatively desirable state of affairs?

In response to the first question, there is a family of initiatives going under the banner of "restorative justice" that appear to be doing just that.¹⁶ Although restorative justice practices take a variety of forms, what makes them distinctive is a nonadversarial form of sentencing in which the various parties or "stakeholders" involved in or otherwise affected by the crime are brought together in a conference.¹⁷ This includes the offender, the victim, supporters of each the offender and victim (often family members or friends), and often other representatives of the community affected by the crime. The formal purpose of the conference is to discuss the effects of the crime (in particular the nature of the harms done) and what can be done by the offender (materially and symbolically) to repair or make amends for such harms. And while the

¹⁴ This dialogical conception of blame amounts to denying the common philosophical thought that blame is essentially something that individuals feel and do, as it may be in the privacy of their own hearts. On the dialogical conception, it is a constitutive feature of blame that it *aims* to engage the wrongdoer in a certain kind of conversational exchange, a fact that explains why blame is characteristically (though not invariably) enacted in social space.

¹⁵ For an importantly nuanced discussion of the complexity of moral repair, see Walker (2006).

¹⁶ While there is a vast and growing literature on restorative justice, good introductory discussions can be found in the following: Johnstone and Van Ness (2007), Menkel-Meadow (2007), Zehr (2002), Van Ness and Strong (2010), Braithwaite (2002).

¹⁷ The guilt of the offender is typically established before a restorative process begins, and sometimes restorative practices are used in a mitigating way after standard sentencing is under way.

offender is expected to take responsibility for his or her offense, in part by coming to a deeper understanding of how his or her behavior has affected others through directly encountering their personal accounts of how they experienced the crime, the focus on repair is meant to serve the larger (forward-looking) goal of establishing or reestablishing a sense of shared community—encompassing, in particular, the victim and the offender, but in which all participants feel they have a stake in upholding communal norms.

Empirical research has yet to establish with full certainty the effectiveness of restorative practices across various dimensions. But initial studies are promising with respect to two significant measures, as compared with outcomes from more traditional trial and sentencing procedures: (1) greater offender and victim satisfaction from the perspective of perceived fair-dealing—victims in particular had a greater sense that justice had been done; and (2) a reduction in recidivism rates, at least for crimes not associated with addiction.¹⁸ From the perspective of this paper, these results are not surprising insofar as restorative practices support a psychologically meaningful trajectory of reactive exchanges such as discussed in section 3. In particular the dynamics of a restorative conference are such that the victim is given the opportunity of expressing how the crime has affected him or her (psychologically as well as materially); the offender, likewise, of explaining how he or she came to commit such a crime. In the course of such exchanges, the offender often comes to recognize the harm caused to the victim in a new way, and the victim comes to appreciate that recognition, often because of the contrition the offender naturally displays.

In sum, restorative justice practices reflect central points that have been argued for in this paper. Under restorative justice initiatives, offenders are not treated in a crude consequentialist way as nonagential targets of rehabilitation and/or deterrence; they are assumed to be fully responsible agents that are appropriate targets of moral address and thus appropriate targets of our blaming acts and attitudes. Yet those acts and attitudes are given a more constructive outlet than merely ensuring that the offender receives appropriate payback (in the form of punishment) for the wrong he or she has willfully done: a goal that is considered normatively appropriate on a more retributive model of criminal justice. Instead offenders are encouraged to *take* responsibility for their wrongdoing by coming to see themselves not only as agents of crime but, more important, as agents both of restitution and of recommitment to the standards of moral community. While such restitution may be legally

¹⁸ One of the more extensive empirical studies to date is the Re-integrative Shaming Experiment (RISE) conducted in Canberra, Australia, from 1995 to 2000 (discussed in Strang 2002; see also Ahmed, Harris, Braithwaite, and Braithwaite 2001; Sherman, Strang, and Woods 2000). Further corroborative work is discussed in Sherman and Strang (2007), Robinson and Shapland (2008), and Shapland, Robinson and Sorsby (2011).

required and enforced, what cannot be required are the reactive emotions that victims, offenders, and other stakeholders often feel in the context of a restorative justice conference as they try to come to grips with their own experiences of the crime and the meaning it should have in their shared community. And yet it is this reactive dynamic—in particular, specific trajectories of reactive exchange—that seems to correlate most nearly with genuine recommitment to the standards of moral community, at least as this recommitment is measured in terms of recidivism rates as well as a general feeling among all the stakeholders that justice has been served.¹⁹

I turn now to my second question: Even if restorative practices go some way to capitalizing on the dialogical aspects of blame, is this really a normatively desirable initiative to be advocating in the field of criminal justice? Or are there further normative demands deemed appropriate in this domain that restorative practices cannot address? A fully developed answer to this question must survey in depth what sort of demands are articulated by the two dominant traditions in criminal justice: namely, (as mentioned above), the consequentialist tradition (broadly associated with welfare-maximizing utilitarianism) and the deontological tradition (broadly associated with retributivism). The question, then, will be to what extent restorative practices meet such demands and, where restorative practices fall short, whether or not the relevant demands are in fact normatively misguided. I cannot hope to undertake such a survey here, but I close by indicating, in a highly abbreviated way, how I think such a discussion might go.

To begin with the consequentialist tradition, I assume there would be broad support for restorative practices so far as they actually do a better job than current trial and sentencing procedures in delivering on what, in their view, the justice system should be aiming at: protection of the community and the prevention of crime through rehabilitation and deterrence. Moreover, so far as such practices promote morally valuable goods such as individual moral capacity, dignity, and autonomy, they should only prove more attractive than traditional responses to crime. Indeed, while it is a standard criticism of consequentialist theories of punishment that advocates are forced to make trade-offs in the promotion of various moral goods (e.g., between successful deterrence and respecting the offender's moral autonomy), embracing restorative practices as a preferred response to crime may provide a way of silencing such criticisms.

But what about deontological retributivists? To what extent would they endorse restorative practices as a normatively fitting or even desirable initiative in criminal justice? It might seem as if they would have a normatively compelling objection. After all, in the retributivist view, the primary aim of the criminal

¹⁹ Cf. Braithwaite and Braithwaite, chapter 4, in Ahmed et al., (2001).

justice system should be one of delivering just deserts (or moral payback) to the offender for his or her crime, and this is what punishment—and only punishment—serves to do (see, e.g., Duff 2001). Of course, rehabilitation or deterrence may be a happy side-effect of society's punishment practices, but no such practice should be promoted simply in virtue of furthering such consequentialist desiderata. Is there anything more to be said from the perspective of this paper against the retributive position?

I have two short responses, to be developed at length elsewhere. The first response is intended to cast a psychological shadow over the bright intuitive light that shines as a beacon in the retributivists' normative field of vision. If what I have said about blame is correct—that the psychology is built on a primitive foundation of punitive payback—it would not be surprising to find considerable intuitive support for the retributivist position. And so far as human intuitions are held to be a central and important guide to the design of a normatively appropriate criminal justice system, we can expect there to be a powerful pull toward some kind of retributive system. But, of course, the question remains: Should such intuitions be accorded such normative importance, especially if we can trace their origin to the more primitive features of our social psychology? The answer is not straightforwardly no, but nor is it straightforwardly yes, as retributivists must surely hold.

My second response is intended to cast doubt on a central normative plank in the retributivist approach to criminal justice. According to its proponents, while punishment is considered the normatively appropriate response to crime, the whole focus of retributive theory is on giving the offender his or her just deserts—no more and no less. Hence one of the main justifications of retributivism as a normatively appropriate response to crime is that it provides a *built-in* check on the amount and kind of punishment inflicted on an offender—specifically that such punishment be proportionate to the crime. According to retributivists, this built-in check is entirely lacking in other approaches to criminal justice, and that is a major normative strike against them. However, if what I have said about the moral psychology of blame is correct, this apparent advantage of retributive theory is deeply illusory.

Here is why. On the account I offer, though blame is built on a primitive foundation of punitive payback, it has certain sophisticated dialogical features that simply will not be satisfied by way of satisfying the more punitive aspects of blame. However, if these more sophisticated dialogical features go unsatisfied, then a blamer's blame is likely to persevere even in the face of delivered punishment, supporting a persisting conviction that justice has yet to be done. But now, if justice can be served only by delivering punishment or hard treatment, as retributivists aver, the only viable option will seem to be one of calling for more severe sentencing practices. In other words, the retributive call for delivering just deserts, so far from ensuring proportionality in response to crime, may, under certain circumstances, produce a powerful ratcheting-up

effect in people's estimation of how much punishment is in fact deserved.²⁰ Hence ironically it is retributive theory that is most in danger of breaching the moral constraint of proportionality in criminal justice—a danger it courts by failing to come to grips with the complex nature of human moral psychology. So far as this argument has merit (deriving support, as needs be, from various empirical studies),²¹ it points to an important moral—and the takeaway message of this paper: No normative theory worth its salt can afford to ignore the real empirical nature of the sort of beings to which it is meant to apply.

²⁰ In support of this point, it is worth noting that, especially in the United States, there has been a dramatic upswing in sentencing severity in the past thirty years or so—an upswing that is concomitant with retributive theory regaining ascendancy as an intellectually respectable doctrine in both legal and policy contexts. No doubt there are many causal factors that explain this increase in sentencing severity (for discussion, see Lacey 2008, ch. 2). However, despite its much-vaunted enshrinement of a just deserts moral framework, it is no surprise that this upswing has been in no way checked by the promulgation of retributive principles. In part this is due to a feature of retributive theory noted even by its defenders: that, even though the principle of delivering just deserts ensures ordinal proportionality in response to different types of crime, there is no “natural” proportion between crime and punishment that serves to anchor the cardinality of the sentencing scale (i.e., there is no internal criterion by which to set an upper and lower bound of the appropriate amount of punishment delivered) (Von Hirsch 1986). Still, this feature alone does not explain why politicians’ call to be tougher on crime plays so well in certain democratic arenas and translates so immediately into stiffer sentencing practices. Again, as Lacey argues, there will be a constellation of factors that account for this phenomenon. I merely suggest that the peculiarity of our moral psychology plays no small role in creating the perfect storm conditions that lead to devastating outcomes like the “three strikes” law in California.

²¹ In partial support of this point, I cite a fascinating psychological study by Carlsmith, Wilson, and Gilbert (2008) investigating the affective consequence of punishment on those who inflict it. Using a free-rider paradigm, subjects of the study were encouraged to play cooperatively with three other players in an investment game in order to achieve an optimal group outcome. Nevertheless, players could improve on this monetary outcome by individually defecting while others cooperate (i.e., by free riding). Almost all subjects played cooperatively. (Defecting subjects were not considered in the final experimental result.) Unbeknown to the subjects involved, the three other players with whom each of them played were actually computers: two of these were programmed to play cooperatively (tit for tat); the third was programmed to free ride. Predictably, the free rider did significantly better than the other players.

At the end of the game, subjects were randomly assigned to one of three groups. One group (the punishment group) was given the opportunity to punish the free rider monetarily by using some of their own earnings; other groups were not. Almost all subjects in the punishment group chose to punish the free rider. Subjects in a so-called forecaster group were not given the opportunity to punish but were asked instead to predict how they would feel had they been given the opportunity. Subjects in a third control group just went on to the next step, a ten-minute filler task that was completed by all three groups. After completing the filler task, all subjects were asked to rate their subjective affective state (negative or positive on a 7-point scale) and how much they were still thinking about the free rider (1–7 rumination scale). Here is the startling result: punishers both felt significantly worse and were ruminating more about the free rider than subjects in either of the nonpunishment groups. In other words, punishers did not experience the sense of psychological closure they (or the forecasters) anticipated punishment would bring. The question is why not?

There are a number of possible explanations for this result (see, e.g., the discussion in Carlsmith et al. 2008), but it is certainly congenial with my own account of angry blame. At its most elemental level, blame motivates the desire to strike back at wrongdoers. But at a more sophisticated level, blame aims at changing or regulating the wrongdoers’ attitudes. Hence blamers will not be satisfied at the simple infliction of punishment—there is business left undone. Thus I contend that even though psychological closure may come with a sense of justice being served, we do not really feel that justice has been served unless and until wrongdoers are brought to a recognition of their wrong and, indeed, a change of attitude. More empirical work needs to be done to determine if this hypothesis is correct.

{ 10 }

Free Will Skepticism, Blame, and Obligation

Derk Pereboom

1. Free Will Skepticism and Blame

A central concern in the historical free will debate is whether the sort of free will required for moral responsibility of a central kind is compatible with the causal determination of our actions by factors beyond our control. Since Hume ([1739] 1978) this concern has prominently been extended to whether this sort of free will is compatible with indeterminacy in action. On the skeptical view I endorse, free will, characterized in this way, is incompatible with this type of causal determination but also with the kind of indeterminacy of action that Hume envisioned (Pereboom 1995, 2001, 2007). It is crucial, however, to recognize that the term ‘moral responsibility’ is used in a variety of ways and that the type of free will or control required for moral responsibility in some of these senses is uncontroversially compatible with the causal determination of action by factors beyond our control, which have therefore not been at issue in the debate between compatibilists and incompatibilists. It is potentially open to the free will skeptic to affirm that we are morally responsible in these other senses.

The particular variety of moral responsibility that has been at issue in the historical debate is set apart by the notion of *basic desert*. For an agent to be morally responsible for an action in this sense is for it to be hers in such a way that she would deserve to be the recipient of an expression of moral resentment or indignation if she understood that it was morally wrong, and she would deserve to be the recipient of an expression of gratitude or praise if she understood that it was morally exemplary. The desert at issue here is basic in the sense that the agent, to be morally responsible, would deserve to be the recipient of the expression of such an attitude just because she has performed the action, given sensitivity to its moral status, not, for example, by virtue of consequentialist or contractualist considerations. One might be the recipient of the expression of such an attitude when this expression is either covert or

What is it to blame someone, and when are would-be blamers in a position to do so? What function does blame serve in our lives, and is there a valuable way of relating to one another? The essays in this volume explore answers to these and related questions.

CONTRIBUTORS

Macalester Bell

Christopher Bennett

D. Justin Coates

Christopher Evan Franklin

Erin I. Kelly

Coleen Macnamara

Victoria McGeer

Michael McKenna

Derk Pereboom

T. M. Scanlon

George Sher

David Shoemaker

Angela M. Smith

Neal A. Tognazzini

R. Jay Wallace

Gary Watson

About the Editors

JUSTIN COATES works on issues in ethics, moral psychology, and the philosophy of law. He is the Law and Philosophy Fellow at the University of Chicago Law School.

NEAL A. TOGNAZZINI works at the intersection of metaphysics and ethics on issues of agency, free will, and moral responsibility. His publications have appeared in *Erkenntnis*, *Philosophy and Phenomenological Research*, *The Philosophical Quarterly*, and *Philosophy and Public Affairs*, among other venues. He is an Assistant Professor at The College of William & Mary.

OXFORD
UNIVERSITY PRESS

www.oup.com

Image: Frank Gonzales. "Nostalgia." www.frankgonzales.net.

ISBN 978-0-19-986084-5

9 780199 860845

COATES
TOGNAZZINI

BLAME

OXFORD

BLAME

Its Nature and Norms

Edited by **D. JUSTIN COATES** and
NEAL A. TOGNAZZINI

Blame

ITS NATURE AND NORMS

Edited by D. Justin Coates
and
Neal A. Tognazzini

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press in
the UK and certain other countries.

Published in the United States of America by
Oxford University Press
198 Madison Avenue, New York, NY 10016

© Oxford University Press 2013

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means, without
the prior permission in writing of Oxford University Press, or as expressly permitted by law,
by license, or under terms agreed with the appropriate reproduction rights organization.
Inquiries concerning reproduction outside the scope of the above should be sent to the
Rights Department, Oxford University Press, at the address above.

You must not circulate this work in any other form
and you must impose this same condition on any acquirer.

Library of Congress Cataloging-in-Publication Data
Blame : its nature and norms / edited by D. Justin Coates & Neal A. Tognazzini.
p. cm.

Includes bibliographical references (p.).

ISBN 978-0-19-986084-5 (pbk. : alk. paper)—ISBN 978-0-19-986082-1 (hardcover : alk. paper)

1. Blame. 2. Faultfinding. 3. Criticism, Personal. I. Coates, D. Justin. II. Tognazzini, Neal A.

BJ1535.F3B53 2013

128'.4—dc23 2012014063

1 3 5 7 9 8 6 4 2

Printed in the United States of America
on acid-free paper

For Stephanie and Anna,
wholly praiseworthy partners

{ CONTENTS }

Acknowledgments	ix
Contributors	xi
 1. The Contours of Blame	 3
D. JUSTIN COATES & NEAL A. TOGNAZZINI	
 2. Moral Blame and Moral Protest	 27
ANGELA M. SMITH	
 3. Wrongdoing and Relationships: The Problem of the Stranger	 49
GEORGE SHER	
 4. The Expressive Function of Blame	 66
CHRISTOPHER BENNETT	
 5. Interpreting Blame	 84
T. M. SCANLON	
 6. Blame and Punishment	 100
DAVID SHOEMAKER	
 7. Directed Blame and Conversation	 119
MICHAEL MCKENNA	
 8. Taking Demands Out of Blame	 141
COLEEN MACNAMARA	
 9. Civilizing Blame	 162
VICTORIA MCGEER	
 10. Free Will Skepticism, Blame, and Obligation	 189
DERK PEREBOOM	
 11. Valuing Blame	 207
CHRISTOPHER EVAN FRANKLIN	
 12. Rightness and Responsibility	 224
R. JAY WALLACE	
 13. What Is an Excuse?	 244
ERIN I. KELLY	

14. The Standing to Blame: A Critique	263
MACALESTER BELL	
15. Standing in Judgment	282
GARY WATSON	
References	303
Index	315

{ ACKNOWLEDGMENTS }

First, our sincere thanks to the contributors, both for writing such great essays and also for being such a pleasure to work with.

Second, we are grateful to several teachers and friends from whom we have learned much about blame and moral responsibility over the years, especially John Martin Fischer, Coleen Macnamara, Angie Smith, and Gary Watson. John in particular has been an exemplary teacher, mentor, and good friend to both of us, and he helped us get this project under way.

Third, thanks to the editors and reviewers at Oxford University Press for their patience and insight, especially Peter Ohlin and Lucy Randall.

And finally, thanks to The College of William & Mary, the University of California, Riverside, the American Council of Learned Societies, and Portland State University for the hospitality and financial assistance that made possible the editing of this book and the research for our introductory chapter.

DJC
Riverside, CA

NAT
Williamsburg, VA
April 2012

{ CONTRIBUTORS }

Macalester Bell is an assistant professor of philosophy at Columbia University.
Christopher Bennett is a senior lecturer in philosophy at the University of Sheffield.

D. Justin Coates is the Law and Philosophy Fellow at the University of Chicago Law School.

Christopher Evan Franklin is an assistant professor of philosophy at Marymount University.

Erin I. Kelly is an associate professor of philosophy at Tufts University.

Coleen Macnamara is an assistant professor of philosophy at the University of California, Riverside.

Victoria McGeer is a research scholar at the University Center for Human Values at Princeton University and a fellow in the school of philosophy at the Australian National University.

Michael McKenna is a professor and Keith Lehrer Chair of Philosophy at the University of Arizona.

Derk Pereboom is a professor of philosophy at Cornell University.

T. M. Scanlon is the Alford Professor of Natural Religion, Moral Philosophy, and Civil Polity at Harvard University.

George Sher is the Herbert S. Autrey Professor of Philosophy at Rice University.

David Shoemaker is an associate professor in the Department of Philosophy and the Murphy Institute at Tulane University.

Angela M. Smith is an associate professor of philosophy at Washington & Lee University.

Neal A. Tognazzini is an assistant professor of philosophy at The College of William & Mary.

R. Jay Wallace is a professor of philosophy at the University of California, Berkeley.

Gary Watson is Provost Professor of Philosophy and Law at the University of Southern California.