Blame, deserved guilt, and harms to standing

Gunnar Björnsson, Stockholm University, gunnar.bjornsson@philosophy.su.se

Having unjustifiably and inexcusably wronged another person and blaming ourselves for it, we characteristically feel guilt. We feel bad in a way prompting us to accept the justified anger and judgment of the victim and to set things right: to apologize, mitigate harm, mend the relationship. Moreover, trying to set things right strikes us as the only unproblematic way to alleviate the pain—unless the suffering is disproportional to the seriousness of what we have done, and needs to be turned down a notch. When proportional, the feeling of guilt is, we sense, *deserved*. Mirroring all this is the characteristic angry blame directed at us from the victims or those reacting on their behalf, implicitly or explicitly calling on us to engage in self-blame. When one is guilty, this too seems deserved; when innocent, undeserved.

Reflecting on these phenomena, it is tempting to think that

> BLAME AND DESERVED GUILT: Blaming someone for something presupposes or involves the sense that they deserve to feel guilty about it.[1]

Moreover, our sense that guilt is deserved seems independent of any thought that it would have good consequences. Guilt, we feel, is deserved merely in virtue of what we have done.[2]

Relatedly, it is tempting to think that *being blameworthy* involves or implies deserving blame, that one deserves blame only to the extent that one deserves self-blame, that self-blame is guilt, and that blaming oneself more involves feeling more guilt. From these assumptions, it follows that

> BLAMEWORTHINESS AND DESERVED GUILT: One is blameworthy for something only if one deserves to feel guilty about it.[3] One is worthy of more blame only if one deserves to feel more guilt.

More strongly, one might think that deservedness of guilt, or guiltworthiness, constitutes the explanatory core of blameworthiness.[4] First, an influential Strawsonian tradition understands other-blame as resentment or indignation,[5] and it is natural to see such reactive attitudes as aiming exactly at the target's pained recognition of what they have done.[6] Given this, other-blame seems to presuppose exactly what feeling guilt does: that the target deserves to feel guilt.

---

[1] For the latter, stronger, idea, see Portmore's (forthcoming) suggestion that blame involves its seeming to the blamer that the target "deserves to suffer the unpleasantness of guilt, regret, or remorse".

[2] In the terminology of Pereboom (2001; 2014), the desert thus seems to be "basic".

[3] See e.g. Clarke 2016 about moral blameworthiness. Duggan 2018 defends the closely related thesis that one is morally responsible "in a liability sense for a transgression" only if one deserves to feel guilt.

[4] Carlsson (2017; 2019) defends this though for "accountability blame", which involves reactive attitudes like resentment, indignation and guilt, but not for what he calls "attributability blame".

[5] See e.g. Wallace 1994; Wolf 2011, inspired by Strawson 1962.

[6] See, among many others, Darwall 2006, ch. 4; Macnamara 2015; Rosen 2015; Shoemaker 2015, ch. 3; Fricker 2016.

Second, one might think that it is the implication of deserved guilt that best explains the common idea that blame is a *sanction* that one can deserve more or less of, that degrees of blameworthiness are a matter of the *severity* of the deserved sanction, and that one can therefore only be to blame for things over which one had a high degree of control. This idea might seem to fit badly with an understanding of blameworthiness in terms of other-blame, as the severity of the sanctions imposed by indignation seems highly contingent on the blamed agent's dependence on the blamer's good will and the circumstances in which the indignation is expressed. By contrast, feeling guilt seems to be a form of suffering and so bad for one, and feeling more guilt worse.[7]

These connections between blame, blameworthiness, and guilt are natural and have considerable explanatory potential. But as I will argue, they are fundamentally mistaken. Trying to account for a wider range of cases, we can see that blame does not presuppose that the target deserves to feel guilt and does not essentially aim at the target's suffering in recognition of what they have done, even when taking the form of resentment or indignation. And the blameworthy do not necessarily *deserve* to be blamed, and even when they do, they do not necessarily deserve to suffer.

On the constructive side, I will offer a new explanation of why, in *typical* cases of moral blameworthiness, the agent deserves blame, and in particular the pained recognition of guilt. The explanation will lean on (i) a general account of moral and non-moral blame and blameworthiness and (ii) a version of the popular idea that moral blame in particular targets agents' objectionable quality of will.

1. Non-moral self-blame without guilt or desert

Though most who accept BLAME AND DESERVED GUILT or BLAMEWORTHINESS AND DESERVED GUILT have had moral blame in mind, our focus will first be on cases of non-moral self-blame. The reason for this is two-fold. First, such cases show how some of the aspects that are also central to moral blame bring neither the suffering of paradigmatic feelings of guilt, nor any sense that suffering is deserved. Second, they illustrate the structure that moral and non-moral blame have in common, a structure that will be central to understanding what is special about moral blame and why paradigmatic cases of moral self-blame do involve something that is properly called suffering and is well understood as deserved.

For an example of the relevant sort of non-moral self-blame, contrast three cases where I form a false belief. Suppose first that I form it based on evidence that would mislead anyone. Later learning that the belief is false, I judge that it was in this way epistemically bad, but I don't blame myself for it: that I formed this false belief was merely a matter of bad epistemic luck. Suppose next that I form the belief based on evidence that would mislead nearly everyone, but not those with truly exemplary epistemic habits. Learning that it is false, I also learn that the

---

[7] See Carlsson 2017; 2019. I reject the claim that the control condition of blameworthiness is explained by the sanctioning involved in blame because (i) non-moral blameworthiness can require at least as much control as moral blameworthiness without involving suffering and (ii) the control requirement seems well explained from the assumption that moral blame is grounded in demands that agents care about moral values. (Some of this will be clearer later on; also see Björnsson and Persson 2012; Björnsson 2017a.)

evidence I relied upon had an esoteric statistical property revealing it as unreliable. With my formation of a false belief serving as a concrete lesson, I update my epistemic strategies accordingly, refining my doxastic heuristics. But I don't blame myself for the failure; though I hold myself to high epistemic standards, they don't require knowing about and keeping track of esoteric statistics that I haven't come upon before. Suppose, finally, that I form the belief based on someone else's authority, failing to check who they were and what their evidence was because the belief pleasingly seemed to support a hypothesis I hoped to be true. Upon discovering the falsehood of the belief and realizing why I had formed it, I not only see the episode as a lesson for future belief formation, but blame myself for having formed a false belief because of a substandard responsiveness to the evidence.

My self-blame here is not merely causal self-blame or some weak form of agent regret: in all three cases, my agency was involved in the regrettable formation of the false belief, but only the third occasioned self-blame. Nor is the self-blame merely a matter of pinning the belief on a failure to respond to available reasons, as my belief in the second case was due to my failure to take into account a statistical property of the available evidence. It is crucial that I fell short of standards that I hold myself to and take to have applied to me at the time. This is why the epistemically bad formation of a bad belief highlights something (perhaps mildly) alarming about me. And it is why self-blame of this sort characteristically involves an emotional reaction that not only has the bad action or outcome as its object, but also *targets* oneself, and is naturally expressed in a self-directed reprimand.[8]

To substantiate these claims, I will say more below about the nature of self-blame and blame generally. But pending further detail, we can recognize the difference between my reactions in the three cases. We can also recognize that the third case will characteristically involve a negatively valenced emotional response to the badness of the belief, my gullibility, and the involvement of my gullibility in the belief formation, likely in the form of irritation or anger directed at myself. This response is naturally understood as self-blame, and seems fitting in light of what I have done. Still, it does not seem to constitute even very mild *suffering* of the sort characteristic of guilt, such that it is prima facie non-instrumentally *bad* for me to be in it.[9] It also does not seem to constitute a *sanction*, or a state that I *deserve* to be in. And it does not seem to presuppose or involve the sense that I deserve to suffer. Because of all this, it doesn't much resemble paradigmatic cases of moral guilt.

Take these points in turn, starting with the claim that the negatively valenced state of self-blame doesn't constitute suffering or make me prima facie non-instrumentally worse off.

---

[8] My self-blame thus goes beyond mere *criticism*, understood as attempts to get the agent to understand the reasons not to do what they did, why it was the wrong thing to do, and to change behavior (Jeppsson 2016). It also goes beyond negative evaluative and emotional responses to *attributability*, i.e. to bad aspects of agency that do not necessarily violate standards that the agent is held to (see e.g. Watson 1996; Arpaly and Schroeder 2014; Shoemaker 2015), and more characteristically involves self-directed anger or irritation than shame, and so isn't mere "attributability blame" in the sense of Carlsson 2019. But as I will stress, it also lacks the connection to sanctions and harsh treatment standardly associated with *accountability* blame.

[9] I say *prima facie* non-instrumentally bad. One might think that some suffering is overall non-instrumentally good for someone because it is deserved, but that it is deserved exactly because it is prima facie non-instrumentally bad for them.

Admittedly, realizing that some substandard aspect of oneself has resulted in a bad outcome might be deeply painful, especially when the aspect is resistant to change or central to one's self-conception. But it doesn't have to be that way, and most often isn't. Noticing and being momentarily alarmed by my gullibly acquired misbelief, I commit to be more alert to the quality of certain kinds of evidence, and then move on. While my life might have been made marginally worse by my gullibility and my misbelief, it was not made prima facie non-instrumentally worse by fleetingly being in the negatively valenced state of self-blame. Perhaps the state disrupted other activities, and perhaps dramatically so. Indeed, it might be part of the function of this sort of state to disrupt other ongoing activities by calling on me to immediately notice the role my gullibility played in the formation of a false belief and to commit to be less gullible. But suppose that any negative disruptive consequences were insignificant enough to be compensated for by my decreased gullibility. Then I wouldn't have traded anything of value to not be in that state.

One might perhaps think that negatively valenced phenomenal states are necessarily prima facie bad for one. But that's a mistake. What gives phenomenal states their negative valence is not that they are themselves bad, but that they present something as bad: in this case the false belief, my gullibility, and the role of my gullibility in forming the belief. Moreover, such states do not plausibly constitute even mild suffering as long as one is *at peace* with them; as long as one does not occurrently desire not to be in them. Consider sorrow as an example.[10] While sorrow is negatively valenced and can constitute suffering, sorrow with which one has found peace does not. Likewise for regret, disappointment, frustration, and even pain. While the moments of regret, disappointment, and frustration involved in practicing a skill might constitute suffering, they typically do not, and the same goes for the pains of physical exercise and sports massage.[11]

Because being in a negatively valenced state of self-blame doesn't itself constitute a cost to me, it is also not well understood as a sanction. One might naturally think that it is: It is the response to violated expectations, like legal sanctions, and arguably has as its psychological function to bring me to satisfy a certain standard, analogously to how legal sanctions (on some understandings) have as their function to promote abidance by imposing costs on transgressions. But unlike legal sanctions (so understood), it isn't that the state is bad for me that prompts me to be less gullible, to avoid being in that state. Rather, the state is one of being alerted to how the bad gullibility yielded a bad belief, and of being alarmed by this. Compare Christine Korsgaard's claim that "Pain is not the condition that is a reason to change your condition…. It is your perception that you have a reason to change your condition."[12] Setting aside Korsgaard's claim about pain specifically, I similarly say that my negatively valenced feeling of self-blame is my vivid sense that forming the belief was bad, that being gullible in that way is

---

[10] Duggan 2018 interestingly understands guilt as sorrow over what one has done, a kind of sorrow deserved by those "morally responsible in a liability sense for a transgression."

[11] For axiological purposes, one might want to understand pain as an occurrent experiential state that one occurrently prefers not to be in for its own sake. Given that notion, pain might perhaps always be prima facie non-instrumentally bad for one, but then some states that are naturally described as painful fail to be pains.

[12] See Korsgaard 1996: 148.

unacceptable, and that my gullibility was instrumentally bad in virtue of its role in the formation of the belief. The role of the state is to alert me to reasons to improve, not to *be* such a reason.[13]

Being in this state of self-blame is also not something I *deserve*. It might be natural to mistake it for such a state: it is a negatively valenced phenomenal state, and one that seems fitting in light of my gullible belief formation. But that's too quick. As I have just argued, the state isn't bad for me, and it seems that in order for me to deserve something, it has to be bad for me in some sense, not good or neutral. For example, suppose that I recklessly break something of yours. Then I might incur obligations to let you know and to do my best to mend the thing and our relationship, and it is natural to think that I deserve the hardships involved. But if I am eager to let you know what I have done to get on with the mending, and doing so is a relief rather than a burden, it is odd to say that I deserve letting you know.[14]

Moreover, even if my negatively valenced state of self-blame were bad and so could in principle be deserved, it wouldn't actually be deserved, even on the plausible assumption that it is in some intuitive sense fitting. Here are two ways of seeing this: First, if the state had been deserved in the way that feeling guilt is deserved in central cases of moral blameworthiness, it would be prima facie wrong or bad to decrease its intensity and duration. But it isn't: if I could diminish the state to a mere flicker through meditation, without undermining the lessons learned from the occasion, the diminishing of the state would not be a consideration against meditating. Second, for being in the state to be deserved in the relevant sense, it would have to be good *as a matter of justice*, because of what makes me blameworthy.[15] Now, being pained by my gullible misbelief might perhaps be seen as good in virtue of being an integral part of my caring about truth and knowledge, or an integral part of caring about living up to the standard I violated, analogous to how the pain of grief might be seen as good in virtue of being an integral part of caring about what has been lost.[16] And perhaps painful disappointment or self-scorn might seem "deserved" in the sense that it would correctly reflect that I fell short of standards I hold myself to. But it is hard to see how the degree to which I feel bad about my gullible misbelief is a matter of justice.[17]

Neither my self-blame nor my sense that I am to blame for my gullible misbelief seems to come with the sense that I deserve to be pained by the recognition of what I have done. Nor does my self-blame seem to presuppose that I deserve this. As I have argued, the negatively valenced feeling involved is not itself prima facie non-instrumentally bad, and there is no sign that it aims or involves any disposition to impose suffering. While self-directed anger might be involved in this sort of self-blame—one might "feel like kicking oneself"—that anger craves no suffering in response, only that one starts taking the standard with sufficient seriousness.[18]

---

[13] Compare Bok 1998: 167–70.

[14] See Nelkin 2013: 121; 2016: 177; Macnamara 2020.

[15] See Macnamara 2015; Clarke 2016.

[16] See McKenna 2019; though see Nelkin 2019 for criticism.

[17] Compare Macnamara 2020 about attempts to establish claims like BLAMEWORTHINESS AND DESERVED GUILT on the basis of claims about goodness or fittingness.

[18] Portmore (forthcoming) points to psychological experiments confirming the presence of desert-related characteristics in central cases of moral blame, self-blame, and guilt. He further takes these experiments to suggest that the desert element is present also in non-moral blame, based on data indicating strong similarities

Finally, and for all the above reasons, it should be clear that the sort of self-blame at work in many non-moral cases doesn't constitute guilt. I might be annoyed that I formed the belief and might give myself an angry reprimand before moving on. But the negatively valenced state I'm in lacks characteristic elements of guilt: the state isn't prima facie non-instrumentally bad for me, isn't plausibly or intuitively understood as deserved, and involves no sense that I deserve to suffer or that there is anything that has to be set right, apart from the sense that my way of responding to evidence won't do.

Admittedly, some of what I have said about my self-blame might be intelligibly resisted. For suppose that the operative epistemic standards are moral standards, perhaps grounded in the fact that substandard belief formation might harm others or constitute a kind of morally objectionable disrespect of reason or truth. Then it might seem less implausible that I deserve to feel bad about my gullible belief formation. (In section 4, we return to how the nature of the standards involved matters for desert.) All I need here, though, is the idea that some clear cases of non-moral self-blame differ in relevant regards from paradigmatic moral self-blame. Thus consider my case of epistemic self-blame on the assumption that while I violated standards that I justifiably hold myself to, satisfying these standards or holding myself to them is not morally required. Or consider the self-blame of Judy, a formidable juggler. Instead of forming a false belief like me, she miscalculates a throw and loses the rhythm necessary to complete an advanced sequence involving a chainsaw and two torches. Like my gullible belief formation, her sequence was botched because her skill-relevant responsiveness to the situation fell below the exacting standard that she justifiably holds herself to. And like me, she blames herself for the resulting failure, annoyed, before giving the sequence another try and succeeding. What I said about my self-blame goes equally for hers: though her response was substandard, botching the sequence bad, and her state of self-blame fitting, her fleeting negatively valenced state of self-blame didn't make her situation prima facie worse, isn't plausibly seen as a sanction, isn't plausibly something that she deserved in the way guilt seems deserved in paradigmatic cases of moral blameworthiness, and doesn't involve or presuppose the sense that she deserves to suffer.

In light of all this, it seems clear that BLAME AND DESERVED GUILT is mistaken if understood without restriction, as concerned with blame generally. Likewise, BLAMEWORTHINESS AND DESERVED GUILT is mistaken if understood as concerned with blameworthiness generally. For it seems clear that I am to blame for my gullible belief-formation and Judy to blame for her botched sequence, and that we are thus in some sense correct in blaming ourselves for these things and so blameworthy for them (in senses of "correct" and "blameworthy" that I'll return to in section 3).

2. Moral blameworthiness without deserved suffering

Though reflection on non-moral blame and blameworthiness unrestricted versions of BLAME AND DESERVED GUILT and BLAMEWORTHINESS AND DESERVED GUILT, most who have been tempted by claims like these have had restricted kinds of blame and blameworthiness in mind.

---

between guilt and regret of self-inflicted harm (Zeelenberg and Breugelmans 2008). As far as I can tell, however, the similarities in question do not involve what is essential to the relevant sense of desert.

For some, it has been *moral* blame and *moral* blameworthiness. For others, it has been blame that involves *holding someone accountable* and a corresponding relation of *accountability*, i.e. the relation in virtue of which one is liable to be held accountable for something.[19] Eventually, I will argue that even restricted in these ways, the two capitalized claims fail. But pending that argument, the fact that they have to be restricted raises questions:

1. What is it about *moral* or *accountability* blameworthiness specifically that makes blame *deserved*?

2. Why is it that *guilt* specifically is deserved in cases of accountability or moral blameworthiness?

The answer to the first question will be based on an assumption about the standards violated by the morally blameworthy. I understand these as interpersonally valid demands to give people's perspectives and interest and other moral values a certain *weight* in deliberation and action, or to accord them a certain de facto "standing". In violating such standards, one has accorded some value too low a standing compared to one's own perspective or interests. One thus incurs obligations to restore its relative standing, by subordinating one's own perspective or interests to it, and becomes liable to be subordinated by others.

The answer to the second question is that paradigmatic cases of moral blame involve the sense that subordination is required, and specifically a kind of subordination that involves staying with the pained recognition of one's fault.

These answers vindicate the intuitions supporting the capitalized claims. But they also leave open the possibility that one might be morally blameworthy without deserving pain specifically, as well as the possibility of moral blame that works more like non-moral blame. Before spelling out the relevant account of blame and blameworthiness generally and the specifics of moral blame and blameworthiness, consider for illustration:

> *Jiro*: Because his writing has been unusually engaging, Jiro has knowingly ignored a promise to go visit his friend Joan, who is going through a crisis. When he sees that Joan is calling, presumably to hear what is going on, Jiro immediately regrets his priorities. Clearly, it was his insufficient concern for Joan and his exaggerated focus on writing that left Joan hanging at a difficult time. As Jiro has done similar things before, to others, and have become increasingly dissatisfied with this tendency, this realization prompts an immediate value conversion. As a result, he sincerely apologizes to Joan and sets out to make his way there as quickly as possible to make things right. Throughout, he feels a number of emotions. First regret and horror at what he had done and concern for Joan, before the value conversion prompts empathy-driven somber determination to do right by Joan, initially accompanied by relief that he had cast off his mistaken values.

On natural ways of filling in this scenario, it seems that (1) Jiro is morally to blame for not visiting Joan prior to the phone call. (2) Jiro deserves blame for what he has done. (3) There is nothing

---

[19] Carlsson 2017 and Duggan 2018 defend the implication from accountability (or "liability" in Duggan's terminology) to deserved guilt. For the classic piece on the distinction between accountability and neighboring relations between an agent and a possible object of blame, see Watson 1996. Also see n. 8.

that he deserved to feel or do in virtue of what he had done that he didn't feel or do. But (4) neither the explicit details of the story nor what needs to be filled in implies any *suffering* on Jiro's part. It is true that in attending to the fact that he harmed Joan and made his own life prima facie non-instrumentally worse, he reacted with horror and regret. As noted, though, we shouldn't think that being in negatively valenced phenomenal states, including moments of regret or horror, is itself necessarily prima facie non-instrumentally bad for one. And because of Jiro's immediate value conversion, he found himself at peace with the fact that his old values and actions were unacceptable and that he would have to make up for what he had done.

Like our cases of non-moral blame, the case of Jiro suggests that blame and blameworthiness do not presuppose or imply that the target deserves to suffer. Some who have considered cases where the guilty party changes values and does their best to set things right but without suffering the pains of guilt have drawn the opposite conclusion. But they might have imagined the agent, not as Jiro, but as a "wrongdoer who responds to outward blame with a sincere and cheerful promise to do better next time but without a hint of guilt or remorse".[20] Considering such a character, it is natural to think that he "palpably frustrates a desire implicit in resentment"[21] or in any case that he doesn't get what he deserves. But cheerfulness isn't the only alternative to suffering. When we empathically hear someone out who has been harmed, or engage in hard and serious conversation, or humbly consider our faults, or do all these things intermixed, we are not cheerful, and our phenomenal states might be negatively valenced. But we also do not necessarily suffer, or find ourselves in states that are prima facie non-instrumentally bad for us. Jiro doesn't, and still his emotional response seems perfectly adequate to his blameworthiness.

In what follows, I will outline an account of blame and blameworthiness generally and an account of what makes *moral* instances of these special. These accounts, I will argue, can explain why the suffering of guilt is deserved in typical cases of moral blameworthiness but not in cases of non-moral blameworthiness, and not in cases like Jiro's.

3. Blame and blameworthiness

Blame, whether moral or non-moral, has both a *target* and an *object*: *someone* is blamed *for something*. To account for blame, we thus need to understand how it relates to both target and object, what it presupposes about their relation, and what sort of a response it is to these matters.[22]

At a natural level of description, the targets are uniform: agents with relevant capacities.[23] However, objects of blame vary immensely in salient ways, even restricting attention to moral

---

[20] See Rosen 2015: 82–83.

[21] Ibid. 83.

[22] The account here draws on accounts of blameworthiness and blame developed in Björnsson 2017a; 2020a; ms. For reasons of space, I cannot here motivate my preference of this account over its main competitors, but many of these have been designed primarily to handle cases of moral blame, and go wrong or stay silent about cases of non-moral blame that target failures of skill rather than failures of will and involve neither reactive attitudes, relationship modification, protests, or holding the object of blame against the target of blame.

[23] Blame can target both individual agents and groups of agents (see Björnsson 2020b); focus here will be on the former.

blame. We readily blame people for attitudes, decisions, intentional actions and omissions, acts and failures of noticing or remembering something important, along with various outcomes of all these things. As we look at other domains, it seems again that a wide variety of responses or failure to respond to aspects of the circumstances matter, as well as a variety of outcomes of such failures: failure to notice or appreciate an opening in a game of chess, to adequately balance force and direction of the foot hitting the soccer ball, or to balance colors or proportions for a painting, resulting in an avoidable loss, a failed pass, and forgettable art. One thing that objects of blame have in common, though, is that they are understood as bad within the relevant domain, as something to be avoided. Another commonality is that they are due to a substandard responsiveness to aspects of the circumstances: a responsiveness that falls short of the blamer's demands or normative expectations on the agent.

Our concern with this complex of conditions, I have argued, is primarily due to our concern with *outcomes in domains or skill*. A *skill*, in the relevant sense, is an *ability* to respond to circumstances so as to promote certain *values*. Specifically, it is a way of responding to circumstances that has been *improved* in response to successful or failed *promotion* of those values, an improvement that has been possible because the degree of promotion is systematically affected by how one responds to circumstances. Thus, acquiring knowledge or forming true, justified, or excellent belief are skills aimed at promoting epistemic values, as are abilities to notice, seek, and weigh evidence. Juggling is a skill aiming at excellent juggling, as are its component abilities, including various abilities to throw, catch, and track projectiles. Caring about what is morally important or valuable is also a skill, aimed at promoting the objects of caring, as are various aspects of caring well so as to better promote these objects, including dispositions to notice and be motivated by opportunities for promotion.

Blaming requires three interrelated abilities to identify certain kinds of conditions and responding to these so as to build and uphold certain levels of skill:

First, the ability to identify the value of relevant outcomes, required for learning what outcomes to promote or prevent.

Second, the ability to identify cases where good or bad outcomes are due to the responsiveness of agents so that the responsiveness can be evaluated as instrumentally good or bad and reinforced or adjusted accordingly. This involves distinguishing when modes of responsiveness explain outcomes in line with a systematic tendency to do so (in "normal" ways) and when they "just happen" to have these outcomes.

Finally, since we often rely on certain levels of skill, the ability to identify and react to instances falling below such levels so as to promote correction. Differently put: the ability to *demand* or *normatively expect* a certain skill level, in weak senses of "demand" and "normative expect", or to *hold the agent to* a certain standard of skill.

When I say that we have *abilities* to do these things, I mean that we have mechanisms the function of which is to make it the case that we do these things, in the sense that these mechanisms continue to play the role they do in our psychology because they have done this.[24]

---

[24] For the sort of account of etiological function that I have in mind, see Millikan 1984.

The mechanisms in virtue of which we have and exercise these abilities operate prior to intentional guidance, and without the involvement of reflective understanding of what exactly the values or demands are or what makes the explanatory connections relevantly normal. In a non-factive and metaphorical sense, we just *see* these things. Specifically, in jointly exercising these abilities we can come to see something as someone's "fault":

FAULT: For Y to be X's fault is for Y to be bad and explained in a normal way by X's responsiveness falling short of applicable standards.

Seeing something as someone's fault, I claim, is the *core* component of blaming someone for something. Judy immediately saw that the botched sequence was her fault, and Jiro immediately saw that having left Joan waiting was his fault.

Given the functions of the three abilities involved, seeing something as someone's fault plausibly has a more specific function:

FAULT FUNCTION: The specific general function of seeing Y as X's fault is to prompt satisfaction of the relevant standard of responsiveness. The normal way for this to happen is (i) for the "demand" mechanisms to direct negative attention to X's substandard responsiveness, (ii) for the mechanisms for feedback learning to direct negative attention to the role of that responsiveness in bringing about Y, and (iii) for the negative attention to these features to prompt improvement of the responsiveness.

The performance of this function is most straightforward in the case of self-blame, which is our main focus here. But we can be alarmed when the substandard responsiveness of another agent has bad outcomes, taking their case as a lesson for ourselves, or convey the lesson to them as well as to third parties.

We can now spell out a general account of blame:

BLAME: To blame X for Y is for one to see Y as X's fault and for the mechanisms normally performing the function of this state to be operative.[25]

On this account, if one thinks that something was someone's fault without that thought prompting activity in the demand and learning mechanisms, one isn't blaming (other than in a dispositional sense). Still, the mechanisms can be operative without successfully performing their full function.

I have said that blame is guided by our sense that something is someone's fault. Our intuitive sense that blame is fitting or correct in some cases but not others primarily tracks the very same condition.[26] Of course, this sense of fittingness might be guided by unjustifiable values or demands. But idealizing these away, we can say that:

---

[25] The quasi-perceptual state of seeing A as F is distinct from the explicit personal-level judgment that A is F. Indeed, I might accept that something isn't my fault but nevertheless, at some level, see it as my fault and so, at that level, blame myself for it.

   The account concerns the psychological activity of blaming. There is also blaming as a communicative act or a speech act, as when someone says that Y was X's fault. Here my focus will be primarily on the former.
[26] Given FAULT FUNCTION, this is also arguably what blaming someone *represents*, in the sort of naturalistic sense of "represent" developed by Ruth Millikan, beginning with her (1984). That the activity of blaming is

> BLAMEWORTHINESS: X is *to blame* for Y, X is *blameworthy* for Y, and blaming X for Y is *fitting* or *correct*, if, and only if, Y is bad relative to a justifiable value and explained in a normal way by X's responsiveness falling short of applicable justifiable standards.

The account sketched in this section makes sense of blame's complex structure. Blame has an object and a target because it involves both a negative evaluation of the object and normative expectations directed at the agent whose substandard responsiveness explains the object. It also makes sense of blame's pervasiveness, because it relies only on abilities and responses that are central to domains of skill. And it explains why blame is embedded in or accompanied by a wide variety of emotional reactions: disappointment or horror in light of the object of blame, regret of one's involvement in bad outcomes, often overshadowed by arousal in the form of irritation and anger aimed at dealing with the problem, and sometimes shame, disgust, and disdain, as agential failures are implicated. As we will see in the next section, the account can also help us understand what is distinctive about central cases of moral blame.

4. Moral blame, deserved guilt, and harms to standing

Equipped with accounts of blame and blameworthiness, let us return to our questions from Section 2: Why do the *morally* blameworthy characteristically deserve certain negative consequences, and why do they often deserve to *feel guilt*, or to be pained by the recognition of their blameworthiness? The answer, I believe, is complex. It involves the nature of the normative expectations that ground moral blame, the nature of the objects of blame, and the relation between these. Seeing its complexity lets us understand both why the pain of guilt often seems deserved, and why it sometimes does not.

Start with four aspects of the demands at work in the moral domain: First, these demands concern agents' *quality of will*: they are demands that agents care about various values enough in relation to other values. Second, these demands are *categorical*: they are demands that agents care about such things whether they want to or not. Contrast these with some non-moral demands, such as the juggler's demands on her juggling skills, which she might understand as contingent on her own preferences. Third, these demands are at least typically *interpersonal*: others justifiably have normative expectations that agents live up to these demands. Correspondingly, violations of the demands that ground moral blame tend to be violations of demands that not only the agent but others can justifiably make on the agent. Fourth, these demands tend to be *requirements of acceptable cooperation*. The categorical interpersonal demands are thus backed up by potential threats of partial or full alienation. If others see one as falling short, they will often be fully cooperative only after one has displayed commitment to the relevant standards.[27]

These first four features of moral demands partially explain the suffering of the morally guilty. The demands inherent in moral blame are demands that one changes what one cares

---

fitting or correctly represents the world doesn't mean that it is all things considered wise or morally appropriate; it might have bad consequences, or the blamer lack standing to blame (see D'Arms and Jacobson 2000).

[27] For an influential discussion, see Bennett 2002.

about. Doing so involves *abandoning* some things that one cares about, at least to some degree, by downgrading their relative importance. Prior to a completed change, this will be understood as pro tanto bad and will tend to be subject to partial inner resistance. When one is morally blameworthy, accepting these demands will thus tend to hurt.[28] Add to this that one might have been seen by others as failing expectations, and that one's valuable relationships might be threatened, and the pain of abandoning some things one values might be accompanied by social fear. Moreover, the process of coming to accept self-blame often involves coming to see clearly, with engaged empathy, not only how others might see one, but how they have been harmed. In such cases, the empathic mirroring of their pain and the regret involved in seeing what one has brought about can further contribute to suffering.[29]

All this also points to important reasons why moral self-blame triggers attempts to mend relationships and show others that one is committed to the norms one has violated.[30] And it explains why we might have not only reasons but obligations to engage in the processes that bring the suffering. For the following principles seem plausible:

> OBLIGATION FROM HARM: If some undeserved harm (material, psychological, social…) happened because of X, X acquires a pro tanto obligation to help set things right.

> OBLIGATION FROM POSITION: If X is in a unique position to help alleviate some harm, X has a pro tanto obligation to do so.

Given these, if I am morally to blame for harming a relationship, I typically acquire a pro tanto obligation to help mend it: the harm was due to me, and I am uniquely placed to set things right. And setting things right will mean accepting the demands of moral blame, and so endure the suffering involved.

Still, what has been said thus far leaves unexplained the sense that the suffering characteristically involved in feeling guilty would be *deserved*. That sense, we have said, involves the idea that it is good as a matter of justice, in light of what makes the agent blameworthy, that the agent is subjected to it. But neither of the two principles above tracks whether the harm was due to the agent's substandard caring, or was the agent's moral fault: they would yield analogous obligations when someone harms a relationship because of a non-culpable confusion.

Instead, I will suggest, what explains the implication of desert is that morally blameworthy action brings a moral imbalance in the standing moral values, and in particular the standing of moral agents. While I lack space to give the idea a full defense or articulation, my hope is that the following sketch will make it plausible enough to merit further consideration.

As understood here, the standing of something is a matter of the weight given to it in deliberation and action. The standing of some value might be most visible when we feel that it is compromised, as when someone's authority or interests or some project one cares about is ignored or downplayed: perhaps someone's views are not taken seriously in seminars, their

---

[28] Compare Bennett's (2002) suggestion that guilt is self-alienation.

[29] See Shoemaker 2015: 110–11. Emotional responses to the anger of actual victims or representatives of victims might play a role in developing guilt as response to one's own perception that one has done wrong.

[30] For some important aspects of this, see Bennett 2002.

preferences ignored when making plans for a night out, or perhaps the University cuts teaching and research funding while administration grows.

Crucially, moral blameworthiness is unavoidably tied to standing, so understood. The relevant moral demands are demands that we accord various weights, or demands that we care about them to certain degrees, as a Strawsonian tradition has it.[31] On the understanding I prefer, to care about a value in the relevant sense is to be disposed to notice factors promoting or thwarting it and to spend psychological, social or material resources to promote it.[32] Correspondingly, caring *more* about one thing than another involves being disposed to spend more resources promoting it. Given this understanding of moral demands, the morally blameworthy has always accorded less weight than morality demands to some value – to a person's perspective, well-being, or property, say, or to beauty, truthfulness, or human excellence, , say – and this has resulted in something that is bad in relation to that value.[33]

Importantly, when someone's relative standing is unduly lowered, this is typically not just something that was bad at the time. The imbalance in standing is an ongoing harm to that person. Perhaps when we last interacted, I gave less weight to your perspective relative to mine than morality allows. Then it remains true that I am unacceptably giving your perspective less weight than mine until this imbalance has been corrected. Compare: If I took something that was rightfully yours, it was not just something bad that happened back then. Your right to that thing is given less weight than it should until this has been corrected.

What is it to correct an imbalance in standing? Generally speaking, it is to do more to promote the value that has been unjustly set aside or thwarted, giving less weight to that which has been unjustly favored. Concretely, if your interest in some divisible good has been unduly favored over mine, then the most obvious way of correcting the imbalance is for you to favor my interest by handing me some of that good. If my perspective has been unduly favored over yours in discussion, correcting the imbalance might mean giving your perspective more attention and authority while giving mine less. (A morally blameworthy agent has always unjustly favored their own perspective; typically, both perspectives and concrete material or social interests of victims have been unfairly disregarded.)

Because standing is a matter of weight given, correcting many forms of imbalance requires the involvement of a particular weight-giver. If I haven't given enough weight to your interests, then others might compensate for that by giving you a helping hand and me the cold shoulder: the community at large might then have given our interests the right relative weight. But it remains true that *I* have, overall, inappropriately put my interests before yours. Only I can correct for that imbalance. How? Not just by coming to care about your interests to the required degree: it will still remain true that I have, overall, unduly put my interests before yours.

---

[31] See Strawson 1962; McKenna 2012; Arpaly and Schroeder 2014. Shoemaker (2015) takes the quality of will relevant to the sort of blame that concerns us here to be quality of regard for persons; I take persons to be (very important) special cases of moral value.

[32] Björnsson 2017a; b

[33] One might be to blame for harms of moral importance even if one cared as much as morality demands, being sufficiently disposed to notice and act on moral reasons, because one makes some other kind of mistake. But then one isn't *morally* to blame for the harm (Björnsson 2017b: 142–44).

Correcting the imbalance means putting your interest before mine in action. Similar things can be said about other forms of imbalance. If I have unduly failed to give your perspective on things the appropriate attention and weight compared to mine, correcting the imbalance does not just mean becoming more disposed to take you seriously. It means actually taking you more seriously, compared to myself, in thought and action.

Based on this, here is why the morally blameworthy deserve setbacks—why such setbacks are good as a matter of justice—and why they characteristically deserve the suffering involved in feeling guilty:

First, as I understand it, the demand to give a certain weight to the perspectives or interests of others relative to your own is not just a demand to give such weight on each occasion, but a demand to properly balance *over time* the psychological and material resources invested into promoting their perspectives or interests and yours. Moreover, doing so is, in some intuitive sense, good as a matter of justice. From this it follows that if I have given your perspective or interests less weight relative to mine than required, then morality demands, as a matter of justice, that I correct this imbalance.

Second, if the guilty have an interest in avoiding the suffering characteristic of guilt, giving no weight to that interest is one way in which they can lower their own standing relative to the victim. By ignoring this interest in promoting the interests and perspective of the victim, the guilty can thus to some extent discharge the demand to rebalance after having put their own perspective over that of the victim.

This provides the outline of an explanation of why the morally blameworthy deserve to have their interests set back, in the sense that it is to some extent good, as a matter of justice. Moreover, the set-back of interest is deserved, not because of what further consequences this might have, or for contractual reasons. It is deserved exactly because the agent is morally blameworthy.

We also have an explanation of why, in central cases of moral blameworthiness, it is natural to think that the blameworthy deserves the feeling of guilt specifically, or the suffering involved in recognizing what they have done.

Still, what the guilty deserve on this explanation is not necessarily suffering. It is that their standing is lowered relative to the interests and perspectives that they have previously unduly subordinated to their own. This is why, when Jiro takes in his wrongdoing, reacts with horror, and proceeds with somber empathy-driven determination to right his wrong while setting his other interests aside, he doesn't *also* deserve to suffer. Correspondingly, while resentment, indignation, and guilt sometimes prompt action that rebalances standing by inflicting suffering on the blameworthy, rebalancing might also often be achieved by prompting the sort of response displayed by Jiro.

5. Loose ends: reactive attitudes, demands on caring

Running out of space, I have to leave loose ends.

In particular, I have said little about the reactive attitudes that figure centrally in many accounts of moral blame, blameworthiness, and responsibility. Here I can only indicate in slogans what I find plausible given what I have said about blame and blameworthiness: *Guilt* is

the sorrowful sense of having unduly disfavored someone or something[34], *resentment* the angry sense of having been unduly disfavored, and *indignation* the angry sense that someone or something has been unduly disfavored. In each case, the object of concern is an imbalance in standing, and in each the action tendency is towards rebalancing. Because moral blame n particular presupposes that the target has unduly disfavored someone or something, it is naturally accompanied by one of these attitudes. And since the morally blameworthy are uniquely placed to correct the resulting imbalance, targets of moral blame are also naturally targeted by guilt, resentment, or indignation.

I have also provided only the barest outline of how demands to give weight to values over time ground reasons to rebalance in response to blameworthy action, and said nothing about how morality constrains such rebalancing, or about the nature and substance of the demands in question. Still, I hope to have made plausible in this paper not only (i) that blame doesn't in general presuppose that the blameworthy deserves to suffer, but also (ii) that the morally blameworthy might deserve setbacks exactly because they have given themselves undue weight and so harmed the relative standing of others or of other values. If the latter explanation seems promising, the exact nature of demands on caring and rebalancing are worth exploring more closely.[35]

Bibliography

Arpaly, Nomy and Schroeder, Timothy 2014: *In Praise of Desire*. Oxford University Press.

Bennett, Christopher 2002: 'The Varieties of Retributive Experience'. *The Philosophical Quarterly*, 52, pp. 145–63.

Björnsson, Gunnar 2017a: 'Explaining (Away) the Epistemic Condition on Moral Responsibility'. In *Responsibility: The Epistemic Condition*. Robichaud, Philip and Wieland, Jan Willem (eds) New York: Oxford University Press pp. 146–62.

Björnsson, Gunnar 2017b: 'Explaining Away Epistemic Skepticism About Culpability'. In *Oxford Studies in Agency and Responsibility*. Shoemaker, David (ed) Oxford University Press pp. 141–64.

Björnsson, Gunnar 2020a: 'Being Implicated: On the Fittingness of Guilt and Indignation over Outcomes'. *Philosophical Studies*.

Björnsson, Gunnar 2020b: 'Collective Responsibility and Collective Obligations without Collective Agents'. In *The Routledge Handbook of Collective Responsibility*. Bazargan-Forward, Saba and Tollefsen, Deborah (eds) New York: Routledge pp. 127–41.

Björnsson, Gunnar ms: 'Skills, Values, and Demands: A Map of Blame and Its Surroundings'.

Björnsson, Gunnar and Persson, Karl 2012: 'The Explanatory Component of Moral Responsibility'. *Noûs*, 46, pp. 326–54.

Bok, Hilary 1998: *Freedom and Responsibility*. Princeton University Press.

Carlsson, Andreas Brekke 2017: 'Blameworthiness as Deserved Guilt'. *The Journal of Ethics*, 21, pp. 89–115.

---

[34] Or, in an extended sense, of having been unduly favored in relation to someone or something.

Carlsson, Andreas Brekke 2019: 'Shame and Attributability'. In *Oxford Studies in Agency and Responsibility, Vol. 6*. Shoemaker, David (ed) pp. 112–39.

Clarke, Randolph 2016: 'Moral Responsibility, Guilt, and Retributivism'. *The Journal of Ethics*, 20, pp. 121–37.

D'Arms, Justin and Jacobson, Daniel 2000: 'The Moralistic Fallacy: On the 'Appropriateness' of Emotions'. *Philosophy and Phenomenological Research*, 61, pp. 65–90.

Darwall, Stephen L. 2006: *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, Mass.: Harvard University Press.

Duggan, A. P. 2018: 'Moral Responsibility as Guiltworthiness'. *Ethical Theory and Moral Practice*. 291–309.

Fricker, Miranda 2016: 'What's the Point of Blame? A Paradigm Based Explanation'. *Noûs*, 50, pp. 165–83.

Jeppsson, Sofia 2016: 'Accountability, Answerability, and Freedom'. *Social Theory & Practice*, 42, pp. 681–705.

Korsgaard, Christine M. 1996: *The Sources of Normativity*. Cambridge: Cambridge U. P.

Macnamara, Coleen 2015: 'Reactive Attitudes as Communicative Entities'. *Philosophy and Phenomenological Research*, 90, pp. 546–69.

Macnamara, Coleen 2020: 'Guilt, Desert, Fittingness, and the Good'. *The Journal of Ethics*, 24, pp. 449–68.

McKenna, Michael 2012: *Conversation and Responsibility*. New York: Oxford University Press.

McKenna, Michael 2019: 'Basically Deserved Blame and Its Value'. *Journal of Ethics and Social Philosophy*, 15.

Millikan, Ruth Garrett 1984: *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.

Nelkin, Dana Kay 2013: 'Desert, Fairness, and Resentment'. *Philosophical Explorations*, 16, 117-32.

Nelkin, Dana Kay 2016: 'Accountability and Desert'. *The Journal of Ethics*, 20, pp. 173–89.

Nelkin, Dana Kay 2019: 'Guilt, Grief, and the Good'. *Social Philosophy and Policy*, 36, pp. 173–91.

Pereboom, Derk 2001: *Living without Free Will*. Cambridge: Cambridge U.P.

Pereboom, Derk 2014: *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.

Portmore, Douglas W. forthcoming: 'A Comprehensive Account of Blame: Self-Blame, Non-Moral Blame, and Blame for the Non-Agential'. (In this volume.)

Rosen, Gideon 2015: 'The Alethic Conception of Moral Responsibility Gideon Rosen'. In *The Nature of Moral Responsibility: New Essays*. Clarke, Randolph, McKenna, Michael and Smith, Angela M. (eds) New York: Oxford University Press pp. 65–88.

Shoemaker, David 2015: *Responsibility from the Margins*. Oxford University Press.

Strawson, Peter F. 1962: 'Freedom and Resentment'. *Proceedings of the British Academy*, 48, pp. 187–211.

Wallace, R. Jay 1994: *Responsibility and the Moral Sentiments*. Cambridge, Mass.: Harvard U. P.

Watson, Gary 1996: 'Two Faces of Responsibility'. *Philosophical Topics*, 24, pp. 227– 48.

Wolf, Susan 2011: 'Blame, Italian Style'. In *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon*. Wallace, R. J., Kumar, R. and Freeman, S. (eds) Oxford: Oxford University Press pp. 332–47.

Zeelenberg, Marcel and Breugelmans, Seger M 2008: 'The Role of Interpersonal Harm in Distinguishing Regret from Guilt'. *Emotion*, 8, pp. 589–596.