



# Blame for me and Not for Thee: Status Sensitivity and Moral Responsibility

Henry Argetsinger<sup>1</sup>

Accepted: 27 January 2022 / Published online: 4 February 2022  
© The Author(s) 2022

## Abstract

In our day-to-day lives, we form responsibility judgements about one another – but we are imperfect beings, and our judgments can be mistaken. This paper suggests that we get things wrong not merely by chance, but predictably and systematically. In particular, these miscues are common when we are dealing with large gaps in social status and power. That is, when we form judgements about those who are much more or less socially powerful than ourselves, it is increasingly likely that “epistemic distorters” will reliably disrupt our reasoning. In response to this situation, I argue for the introduction of a new norm concerning blame and social status: “Powerful Restraint.” This norm suggests that the socially powerful ought to blame the socially dispossessed less frequently, and that social pressure be applied to achieve that compliance. I argue that such a norm is justified given the pernicious down-stream moral effects of the up-stream epistemic distortion I describe. I defend it against objections that it suggests a normative solution to an epistemic problem, that it is paternalistic, and that it involves a kind of disrespect to the less socially powerful.

**Keywords** Moral Responsibility · Social status · Ethics of blame

In our day to day lives, we form responsibility judgments about each other. We ascribe responsibility, praise, blame and hold one another to account. Sometimes, these attributions of responsibility are wrong – we are imperfect beings, and our judgments can be mistaken. It’s possible, then, to imagine that we might get things wrong predictably and systematically. Imagine that, sometimes, we form responsibility judgements in ways that are disconnected from the facts about capacities, selves, and attitudes that *ought* to matter, according to our best metaphysics of responsibility. That is, where we ought to look for things like “quality of will”<sup>1</sup> or “reasons responsiveness,”<sup>2</sup> we instead make use of affective cues about likeability, or just-so narratives, or vulgar stereotypes. Imagine further that these miscues are particularly common when we are dealing with large gaps in social

<sup>1</sup> See, for instance, Strawson (2003), Scanlon (2008), and McKenna (2012).

<sup>2</sup> See, for instance, Fischer and Ravizza (1998), Brink and Nelkin (2013), and Vargas (2013).

---

✉ Henry Argetsinger  
hargetsi@ucsd.edu

<sup>1</sup> Philosophy Department, University of California, San Diego, 9500 Gilman Dr., 0119, La Jolla, CA 92093-0119, USA

status. That is, when we form judgements about those who are much more or less socially powerful than ourselves, it is increasingly likely that “epistemic distorters” will reliably disrupt our reasoning.

There is good evidence that the imaginary world I’m describing is a more or less accurate picture of how our world actually operates. If the world *is* something like that, we can ask: what should we do? In this paper I propose an answer to that question, arguing for the introduction of a new set of norms concerning blame and social status. These norms suggest that the socially powerful ought to blame those with less power *less* frequently.<sup>3</sup> The goal of the paper, then, is to articulate and defend a norm suggesting that the powerful restrain their blaming tendencies in certain ways.

In Part I. I explain some of the background motivations and assumptions for the view: a) that a fleshed-out epistemology of moral responsibility shows that many of the judgements about responsibility we make do not track the metaphysical properties that responsibility theories say they ought to, and b) that social power, in particular, distorts our judgements of responsibility. In Part II. of the paper I explore one kind of revisionist norm: a blanket reduction of blame across our practices. I argue that this is unlikely to produce ameliorative results for a variety of reasons. In Part III. I argue that an asymmetric set of norms is preferable: in particular that the powerful ought to blame less. Finally, I consider some objections to this view, particularly that asking the powerful to refrain from blaming is objectionably paternalistic or disrespectful to the powerless.

## 1 Background: Epistemic Distortion and Social Status

Although there has been a great deal of excellent philosophical work on the metaphysics of moral responsibility, very little attention has, historically, been paid to a corresponding epistemology of responsibility.<sup>4</sup> This is perfectly understandable: as philosophers we want to know in virtue of what someone *is* responsible. This is, in some sense, a more basic question than how we reason towards, arrive at judgments of, or know about one another’s responsibility.<sup>5</sup>

Yet, the lack of attention to everyday epistemology presents some problems for theorists of responsibility who tacitly assume a kind of “metaphysics first” view of how our epistemic processes *should* work. Say, for instance, you argue that having a reasons-responsive mechanism that operates (in a characteristic range of cases) in an understandable pattern is what’s required for an agent to be responsible.<sup>6</sup> Such a view implies that the *right* way to investigate, come to agreement about, justify, or verify responsibility judgements would be to find out if the agent really *does* have this kind of mechanism (and whether it is operating in a given case). Perhaps this *is* the right theory of an *idealized* epistemology, but, I argue,

<sup>3</sup> A further suggestion is that the socially less-powerful ought to blame the more powerful *more* frequently, but I do not focus on this avenue of argument here.

<sup>4</sup> In what follows I will use the terms “responsibility” and “moral responsibility” interchangeably. I mean to be discussing moral responsibility, although I suspect my claims may apply to other kinds of responsibility as well.

<sup>5</sup> There’s also a basic sociological explanation about path dependency – those in the literature are responding to a metaphysical debate, but we can easily imagine a path not taken where the epistemology of responsibility was central.

<sup>6</sup> As Fischer and Ravizza (1998) do.

this is not what those in our day-to-day, real-world moral responsibility practices actually do.<sup>7</sup>

Because of this, we ought to pay attention to how the real-world epistemology of moral responsibility operates. Once we do, we'll notice that the ideal picture of reasoning (whatever the correct theory of responsibility is), is often "distorted" by factors which are extraneous from the point of view of that metaphysics. This is especially apparent if we are working within a framework of social epistemology focusing on questions about disagreement, ignorance, and epistemic injustice.<sup>8</sup> When we look at the non-ideal realm, in other words, it seems we don't come to responsibility judgements the way the metaphysical theories (tacitly) predict we will.

What explains this? One key finding of the last several decades of work in social and cognitive psychology, particularly in the "attribution theory" and "social comparison literatures, is that when we attribute beliefs, intentions, actions, and traits to agents, we often do so without conscious deliberation, by making use of features of situations (and agents) such as: warmth and competence,<sup>9</sup> social proximity,<sup>10</sup> heuristic stereotypes,<sup>11</sup> facial features and attractiveness,<sup>12</sup> and estimations of causal control.<sup>13</sup> Insofar as we *do* consciously deliberate, we are often focused on things like interpreting the social meaning of an action, making sense of an agent's "character" in the context of how others judge them, or determining how good or bad an outcome was.<sup>14</sup> In either case, these kinds of conscious and non-conscious inputs are all extraneous from the point of view of most metaphysical theories of responsibility. To be clear, I do not mean that something like character or outcome-luck is extraneous in-itself for such theories. The point, instead, is that those theories exhort us to react to the facts that are (or would be) actually constitutive of an agent's character, for instance. In day-to-day life we do not have direct access to these metaphysical features, and are often reacting, instead, to proxies and heuristics that stand in for these facts. A split-second judgment of "warmth and trustworthiness" does not tell me about, for example, your actual quality of will or character traits.<sup>15</sup>

One particular kind of epistemic distortion is that of what I call "status sensitivity." We are often affectively biased, both in our initial reactions, later searches for evidence, and overall interpretation of events, by whether someone has a relatively high (or low) degree

<sup>7</sup> Nor is it what they see themselves as doing - although self-conception isn't always the best guide to the normative ideal of a practice.

<sup>8</sup> See Dotson (2014), Fricker (2007), and Manne (2018) as representative recent examples of this framework.

<sup>9</sup> See for instance: Cuddy et al. (2007), Feigenson (2016), Fiske et al. (2007), Nadler (2012), Nadler and McDonnell (2012), and Rahimi et al. (2016).

<sup>10</sup> See, for instance: Buunk and Mussweiler (2001), Gerber et al. (2018), Suls and Wheeler (2000), and Suls et al. (2002).

<sup>11</sup> Cuddy et al. (2007), Ellison and Munro (2008), Fiske et al. (2002), Mazella and Feingold (1994), Mitchell et al. (2005), Sommers and Ellsworth (2000), Westra (2018a), and Willemsen et al. (2018).

<sup>12</sup> See, for instance: Mazella and Feingold (1994) and Devine and Caughlin (2014).

<sup>13</sup> See, for instance: Alicke et al. (2008) and Alicke et al. (2015).

<sup>14</sup> See, for instance: Bayles (1982), Brewer (1977), Gailey and Frank Falk (2008), Gawronski (2009), Lagrado and Channon (2008), Nadler (2012), Pizarro and Tannenbaum (2012), Shaver (1985), Westra (2018b), and Westra (2019).

<sup>15</sup> I don't mean to stake any claim about whether or not focusing on these kinds of features is *irrational* - we may have evolved for it to be quite rational to pick up on warmth and trustworthiness, for instance.

I simply mean that they are not part of a proper process of conscious deliberation according to our metaphysical theories.

of social status, and whether their social position is favorable or dis-favorable to us in various ways.<sup>16</sup> I use the phrase social status somewhat interchangeably with the phrase “social power.” In both cases I mean to invoke an intuitive combination of our contextual status in social groups, our social abilities, and our material resources. In other words, our power to act, influence, and gain uptake in the social world. This power is partially constitutive of the social world itself, and is deeply relational.<sup>17</sup> Our positionality affects our agency, our options, our relationships, our status and our values – and all of this is mutable and contextual. In some contexts we may be of high social status and have a great deal of social power, and in other contexts we might be of low social status, or even count as oppressed.

No one denies that social power exists, or that we have typical epistemic processes which result in judgments about moral responsibility. What I’m noting is that the two interact in ways that produce outcomes largely absent from the moral responsibility literature.<sup>18</sup> In particular, attention to the epistemology of responsibility helps us make sense of cases where those with a great deal of social power seem to dodge responsibility generally, and, in particular, *blame*.<sup>19</sup>

In this sense, social power is a particularly reliable and pernicious distorter of the epistemology of responsibility. I don’t have space here to fully work out a theory of epistemology, its distortion, or the nature of power and status. I’ve provided some empirical citations and a sketch of the issue that I hope is plausible. For the rest of the paper, I ask the reader to imagine that our world could be as I describe it. It could be the case that our judgements of responsibility are reliably “distorted” by features of social power. If this is the case, a question naturally follows: what should be done about it?

<sup>16</sup> Importantly, status sensitivity errors will almost always involve and be intertwined with the interpretational errors I’ve noted above. Indeed, much of the empirical evidence for these errors overlaps with the kinds of resources cited in footnotes 5-10 above. But see also work in the social comparison literature, for example: Barden et al. (2004), Krueger (2000), and Zhao and Rogalin (2017).

<sup>17</sup> My thinking about social power draws, in large part, on the work of feminist philosophers such as Cudd (2006), Oshana (2018), and Young (1990), as well as recent work by Abizadeh (2021) and Menge (2020).

<sup>18</sup> One interesting historical precursor to my arguments here is the work of Adam Smith. He held that we have an unhealthy obsession with the wealthy and powerful, and an unfortunate inclination to look down on those in destitution. As he says: “Upon this disposition of mankind, to go along with all the passions of the rich and the powerful, is founded the distinction of ranks, and the order of society. Our obsequiousness to our superiors more frequently arises from our admiration for the advantages of their situation, than from any private expectations of benefit from their good-will... Their benefits can extend but to a few; but their fortunes interest almost ever body” Smith (1976, p. 52).

<sup>19</sup> Consider the infamous, recent case of Brock Turner. Turner, a member of the Stanford Swim team, raped an unconscious classmate, was arrested, and faced prosecution. He blamed his actions on, “a culture of drinking, peer pressure and ‘sexual promiscuity.’” His father lamented, “that his son’s life had been ruined for ‘20 min of action,’” and the judge in the case handed down an extremely lenient sentence. (See: <https://www.nytimes.com/2016/06/09/us/brock-turner-blamed-drinking-and-promiscuity-in-sexual-assault-at-stanford.html>). Compare Turner’s treatment to the way the socially “low-status” are magnets for responsibility and blame ascriptions. Note, also, that there is an interesting asymmetry here: it is not as if the high-status similarly “dodge” ascriptions of praise. However, this asymmetry between praise and blame is beyond the scope of this paper.

## 2 Revisionism about Responsibility: Against Blame?

Very roughly, there are three options concerning our distorted practices: retention, elimination, or revision. I'll argue for revision. Even if we are habitually bad responsibility practitioners, getting rid of the practices is unlikely to be beneficial. Of course, much here depends on how comparatively bad scrapping a practice would be, how likely one thinks revisionism is to succeed, and how radical the revision necessary to improve the practice would be. Given the difficulties of revision, another option is straight retention. Perhaps, in the end, retaining our faulty practices is all we can do – especially if they are psychologically ineliminable. However, given that we recognize the practices are, in some sense, faulty, and given that the norms of the practices are, to some degree, up to us, it seems worthwhile to *try* to revise them, to the extent such revision wouldn't make things worse.

But what kind of revision? There is no easy, practice-wide answer to this question. As I've said, social power is highly contextual and non-static. Good luck figuring out who is more "powerful" in many real-world cases. Protests in May of 2020 in Minneapolis provide a good example of this kind of positionality. After the killing of George Floyd, several days of peaceful protests turned into riots, and eventually a police station was burned to the ground. In such a context, would it be fair to say that a protester, burning a police precinct, has more social power than a police officer in the city? There is no clean answer to this kind of question. In some sense, as an arm of state sanctioned violence and law enforcement, it is hard to imagine a protester having more material power than a police officer. Yet, many observers were supportive of the protesters, and it is clear they wielded enormous social import and status. They were also able, in this instance, to physically overcome the police force, who they greatly outnumber. Again, one can draw their own conclusions here – the point is that analyzing power, because it is non-static and highly contextual, isn't likely to deliver easy answers for us about who should hold who to account, or who should refrain from engaging in the practices of responsibility.

On the other hand, extreme, near-edge cases exist. It would be ludicrous to claim that I have more social power than Jeff Bezos – even if there is some way in which social opinion or sentiment is more sympathetic towards me than the richest man in the world. In any case, it's the overall normative structure of our practices where the real action is. What licenses our retention (or abandonment) of blame in various aspects of our responsibility practices is whether doing so promotes valuable social norms. As Cheshire Calhoun (1989) famously argues, in cases of injustice which occur at the level of social practice, "the question of blame becomes not just a question about blameworthiness, but more important a question about our entitlement to use moral reproach as a tool for effecting social change" (389). We ought to, in other words, ask what the aims of our responsibility practices are, and what configurations of the practices would best meet those aims. Then, I think, it is an open question whether something like blaming across large gaps in power is instrumentally justifiable – and if so, when and how.

Given the fact that determining precise power relations is difficult, one very natural candidate for practice revision is the following principle:

**Blanket Blame Reduction:** Given that gaps in social power are prevalent and commonly cause epistemic distortion, we all ought to be more cautious about our judgments of moral responsibility; in particular we all ought to refrain from blaming across large gaps in social power.

Would the introduction of this kind of norm improve our responsibility practices? I think the norm is misguided, and imagines a silver bullet where, unfortunately, none exists. Its approach to a structural problem in our practices is to treat all the individual members of the practice roughly equally. It asks all of us to refrain from blaming as often as we do (especially when we are aware of gaps in power). There are several problems with this approach. First of all, in asking the less powerful to exercise even *further* humility, it both fails to correct the fundamental imbalance of power that gives rise to the most pernicious problems that I've identified, *and* takes away one of the only tools that oppressed peoples have to fight injustice: social sanction. The socially powerful are often given too *many* free passes. The epistemic errors we make tend to put us on a course towards blaming the powerful too little.

Not only this, Blanket Blame Reduction merely reproduces the flawed normative landscape we've been discussing at a higher order. The introduction of this norm would also introduce a "metanorm" that says we ought to blame those who fail to sufficiently reduce their blaming tendencies. But, because the scope of this norm would include those with low social power as well as those with high social power, the same distortors will manifest at this meta-level as well. The powerful, in other words, won't get their fair share of blame for failing to reduce their blaming tendencies, while the less-powerful will get too much.

Finally, one might be worried that a general prohibition on holding one another responsible will only increase cultivated ignorance.<sup>20</sup> It is hard to see how active ignorance which leads to epistemic injustices can be overcome by a system that asks us all to refrain from blaming for *fear* of being ignorant. It seems entirely open that groups who rely on strategic ignorance to, for instance, maintain a dominant social position without guilt or moral reckoning will be able to further entrench that position if they can act without fear of reproach (and if they can react to blame by saying, "you are blaming me across a large gap in social power - you ought to have reduced your blaming tendencies!").

For these reasons, let me suggest a first pass at a more specific norm which retains an asymmetry in blame's acceptability:

**Powerful Restraint:** the socially powerful ought to, in general, refrain from blaming the less powerful in contexts where large gaps in power are prevalent.

In order to explain why we ought to favor Powerful Restraint I need to do three things: First, I need to say a bit more about responsibility and blame. How precise a notion of these concepts do we need for my argument to go through? Is the kind of distortion I'm canvassing a problem for only some theories of responsibility and blame? Or is it meant to generalize? Second, I need to further clarify the scope and grounds of the norm of Powerful Restraint - to whom, exactly, does it apply? And what grounds this *prima facie* obligation?<sup>21</sup> Finally, I need to motivate the idea that forgoing blame in the way it imagines could be socially beneficial at all, given the worries I canvassed above.

## 2.1 Metaphysics, Scope, and Justifying Blame Reduction

Let me tackle these three questions in order. Regarding blame and responsibility, my hope is to be very ecumenical, and to earn that ecumenicism. Whatever the right view

<sup>20</sup> See, for instance Charles Mills (2017) work on what he calls "white ignorance."

<sup>21</sup> Thank you to two anonymous reviewers for pressing me to clarify these important points.

of the metaphysics of responsibility and blame, we can (and should) ask: “what kinds of things are unnoticed distorters or defeaters of our epistemic abilities to accurately, fairly, or usefully track and judge that people are responsible?” Now, it might be objected that, to answer this question, we’d need a full theory of the epistemology of moral responsibility, and that to give *that* theory, we’d need a fully worked out account of moral responsibility itself. That is, whatever the right epistemic story is, it will have to have some connection with the right metaphysical story.

Do we need, then, to venture deep into the weeds of the metaphysics of responsibility? My claim is that we needn’t take up such issues in the present paper.<sup>22</sup> Instead, we can still advance a robust and interesting view, because we can say the following: “here are a set of practices, conditions, and common facts such that, when they obtain, there are distortions, disruptions and impairments to our judgements of whether and when people are morally responsible.” Notice that, for this kind of claim, one needn’t be committed to a particular view in either the metaphysics *or* the epistemology of moral responsibility to see when and how these distortions will crop up, and why such defeaters matter. The point is to try to get clear about where our epistemic practices of moral responsibility break down and what the causes of those break-downs might be. To do that, we merely need a sketch of our current responsibility practices, from which we can think about what’s going to count as a distortion or defeater under a wide range of credible views in epistemology and the metaphysics of responsibility.

In some sense, my arguments might be read as a kind of quietist position regarding the metaphysics of responsibility. The idea is just that, whatever account one favors, the metaphysical details may be misleading and unfruitful in the epistemic realm – we are unlikely to uncover, for instance, the precise neurological underpinnings of responsibility *and* to make such neurophysiological bases epistemically accessible to everyday practitioners of responsibility (that is, to all of us). Until the arrival of “responsibility-meters” which we can surreptitiously point at our friends and family, the metaphysical facts simply can’t play the kind of grounding role that would matter in a robust way for our epistemic work. So, I suggest that we view the metaphysical facts as a kind of black box. The box exists, and it matters. Into it go various psychological, attitudinal, and capacitarian facts from the the natural and social worlds, out of it comes responsibility. But what happens inside the box needn’t concern us, so much as whether we are giving plausible interpretations of the contours of the box itself.

Concerning blame, a similar argument can be run. Again, my aim is to be as theory neutral as possible. Whether blame is a reactive attitude, a cognitive state such as belief, an adjustment to a relationship, or a form of conversational protest doesn’t matter for the purposes of my argument. We are all aware of the social reality of blame, of the way it feels (both to blame and to be blamed), and of the many forms it can take. Whether there is a univocal (or pluralist) account that we can give of blame’s necessary and sufficient conditions won’t matter for the two points I make in this paper. First, that there is an epistemic problem in our practices of moral responsibility, and second, that this problem leads to our blaming badly, and our doing so reliably. Insofar as I take a stand on the many interesting questions of the nature of blame it is only to say the following: however one wishes to precisely explore the contours of blame as an emotional, cognitive, or interactional process, we should be able to recognize that blame has what I’ll call an “assessment phase” and

<sup>22</sup> For those who are unconvinced, note that I have also treated these issues at much greater length in other work: Henry Argetsinger (2021), “Blaming Badly,” Manuscript submitted for publication, and Henry Argetsinger (2022), *The Epistemology of Moral Responsibility*, PhD diss., (UC San Diego).

an “expression phase.” The process of coming to form a judgment of blameworthiness is separate from choosing how we express (or do not express) that judgment to ourselves, the blameworthy individual, or others. Different theories treat these phases in different ways, and I take no stance on the correct way of thinking about the distinction between blameworthiness and blame - all I need to point out is that a gap between judgment and action exists here.

Next, let me clarify the scope and grounding of the norm of Powerful Restraint. In section one, I canvassed a set of epistemic issues that led me to claim that it’s likely that the powerful often blame the less-powerful too frequently. Given this, one might ask of Powerful Restraint: Isn’t the norm unnecessarily strong?<sup>23</sup> That is, why think that anyone needs to, in general, reduce their blaming tendencies? Wouldn’t a more judicious norm ask them to *raise* the epistemic bar their blame needs to clear, rather than reduce it *tout court*? We can imagine that if the powerful focused on better collection of evidence, on correcting for biases, and on double-checking their accounts of what agents did or why they seem blameworthy, the tendency for blame to be misapplied across gaps in power would be reduced. There are two responses I can give here. The first makes clearer the relationship between the epistemic problem I canvassed in section one and the moral problems that result from it.<sup>24</sup> The second points out that this raising of the epistemic bar is meant to be contained in the structure of the norm itself.

The first answer is that I suspect there is no clean and sharp distinction between the moral and epistemic in the kind of cases I’m considering, and that trying to make a clean distinction will be less helpful than it might at first seem. Nothing I say here precludes the possibility that distinguishing carefully between the downstream moral effects of an upstream epistemic problem is the right way to frame the issue. However, my claim is that these two issues are reliably blended in the case of the epistemology of moral responsibility, and that there’s a specific structure to that blending that is worth dealing with in itself. How so? We have a set of moral practices which ask us to attend to certain kinds and sets of evidence – our best theories of the metaphysics of responsibility tell us to be responsive and attentive to that evidence, and our best theories of epistemology tell us how we ought to gather and assess it.

However, I’ve claimed that it’s both the case that: 1) actual agents in the practice do not follow these standards reliably well, and 2) the standards themselves may be suspect or faulty in various ways. This calls for revisions in the practice, as I’ve argued. And, what’s going to govern the revision will be structured both by general features of good epistemic practice, and by the particular role those epistemic practices play in the practice of moral responsibility. We have, in other words, an overlapping structure – there are independent epistemic norms about how to treat evidence and form reliable judgments and there are internal, moral norms about how to be a good participant in the practices of moral responsibility. The epistemic norms can be seen as necessary but non-sufficient conditions on our forming good judgments of responsibility. But there are also distinct necessary conditions which emerge from the moral responsibility practice itself. To know whether I ought to blame you I need to know that I’m tracking the right kind of evidence (and tracking it well), but I also need to know what my blame will *do*; whether it will cause undue harm, whether it will be *fair*, and so on.

<sup>23</sup> Thank you to an anonymous reviewer for pushing me to clarify this point.

<sup>24</sup> Again, my sincere thanks to an anonymous reviewer for pointing out that the distinction between whether this is a moral or epistemic matter must be made clearer.



Where epistemic problems are likely to occur in our practices, I'm claiming then, moral problems follow close behind. And again, I've argued that these moral problems are not random, but are reliably structured by the practices themselves. If it's the case that the powerful are reliably bad judges of the character of powerless members of society, for instance, then they are reliably likely to blame them unjustly. This leads to a distributional problem – an unjust balance of blame at the societal level. This is a moral issue, but one rooted in a particular epistemic problem. So, Powerful Restraint is moralized - but not haphazardly. It suggests a solution that is partially epistemic and partially moral because the problem it responds to is a downstream moral problem resulting from an upstream epistemic one.

Still, one can press: why not just deal with the upstream problem in isolation? Here the second line of response is called for. The “in general” clause in Powerful Restraint is meant to indicate that there is no blanket prohibition on blame from the powerful - rather there is normative pressure for more care than they often exhibit before blaming, *as well as* pressure for a general reduction of their blaming tendencies (or, an increase in their hesitancy to blame). In other words, the spirit of this objection is already contained in the formulation of Powerful Restraint.

More importantly, however, I am skeptical of the idea that the biases that may lead to epistemic distortion in these cases can be sufficiently corrected or accounted for upstream of blame, such that blame would once again be (in a general way) appropriate. Instead, one thing the paper argues for is that since we, in general, cannot know the precise factors our blame judgments are formed by and react to, and since we cannot know the precise features of agents' metaphysical or characterological make-ups that would vouchsafe blame, we cannot make our blame legitimate *merely* by being more epistemically cautious or working to reduce our biases.

Roughly then, we can say that: a) yes, the powerful (and others) should work to improve their epistemic processes as much as possible - we want to get blame right when we can. But, b) no matter how hard we work, mistakes are possible, and become increasingly likely in the kinds of cases I describe where we operate across large gaps in social power and prestige. Given the moral costs of these mistakes and our inability to eliminate them, I claim that Powerful Restraint is justified.

Finally, why should we believe that forgoing blame would be socially beneficial at all? For two reasons: First, by focusing on the *downstream* products of responsibility judgements, Powerful Restraint is far less revisionary than a view that asks us to do without appraisal in the first place. Recall the worry above that our responsibility practices are psychologically ineliminable. While it's true that our “hot” and non-conscious psychological systems may already be priming us to form judgements before we have a chance to assess evidence or do a great deal of careful cognition, we still often have the chance to reflect before *expressing* those judgements. Forgoing blame, in other words, at least in its *social* or dialogical guises, is far more under our voluntary control than forgoing attributions of responsibility in total.

Second, there is good evidence that forgoing blame can have instrumentally beneficial effects in many important contexts. The work of Hanna Pickard (2013), for instance, argues that such a bifurcation between judgements of responsibility and blame is crucial in clinical psychiatric work. Pickard (2013) notes a uniquely *clinical* conundrum: in institutional settings where service users suffer from disorders of agency (bi-polar disorder, for example), caregivers must hold them responsible for their actions while avoiding blame. This is so because holding responsible is crucial for a) treatment, and b) respecting service users as agents and persons, while blame, on the other hand, is highly detrimental for treatment (see 1135-1138).

The key point is that, although many services users may have *diminished* amounts of control or conscious awareness (or, we can assume, whatever other properties, capacities and faculties one's theory of responsibility calls for), on any notion where these capacities are graded, most will pass a threshold of responsible agency most of the time. Their excuses for diminished responsibility do not exempt them from the practice wholesale. Exemption, perhaps by taking a Strawsonian objective attitude, would deny them their agency – something which would both be disrespectful and counterproductive to the intended therapeutic interventions.<sup>25</sup>

Yet, it's also the case that, in terms of effective treatment, blame, expressed with a characteristic emotional “sting” is highly detrimental in clinical settings. Pickard thus gives us a clear example of a setting where forward looking concerns (the goal of proper and effective treatment) shape the way in which the practice of responsibility takes place in a particular social arrangement. Pickard notes that, quite obviously, the large gaps in power between clinician and patient also play a role in making affective blame ineffective. We are offered a practical, instrumentally justified solution to the problem: take responsibility for our own emotions and affective responses, keeping in mind the complicated power dynamics between patient and caregiver.

All of this is to say that holding responsible looks different in different contexts. Insofar as different contexts call for different instrumental justifications of blame, our responsibility practices can involve revisions to the frequency and type of blame we engage in. The revisions to our practices which are licensed depend on what we think those practices are good for, and how we think we can best achieve those goods. When we look at the context of the high-status blaming the low-status, what further reasons do we have for thinking that reducing the expression of blame may be beneficial?

## 2.2 Issues of Incentivization and Positionality

One reason to minimize the flow of blame from high to low power individuals is due to the following concern: those who are low-status are going to be incentivized towards certain instances of “blameworthy” behavior in a way that those of high status are not. Not only this, but those with power and privilege are often partially responsible for the structure of those incentives to begin with, and so may not be in an appropriate position to blame those of low social-status and power. That is, where large gaps in social power occur, it may be that we are in the wrong kind of relationship to hold one another responsible.

How might this argument work? Lewis (2016), has argued that those who are complicit in creating the conditions which lead to blameworthy behavior, do something inappropriate when they blame. This is a rather intuitive idea: I shouldn't blame you for doing something I enabled (and perhaps foresaw as a likely outcome) - or, at the very least, my blame ought to be limited or tempered. Lewis advances this argument in a context which works well for our purposes: the fact that those who commit crimes are often disadvantaged persons in disadvantaged communities who are strongly incentivized to do so.

As Lewis notes: “Because blame is a response to a perception of a morally inappropriate attitude, it might be natural to think that blame is justified when that perception is accurate. But it is also natural to think that there is an important sense in which our actions and

<sup>25</sup> The objective attitude is famously discussed by Strawson (2003, pp. 79–83).

attitudes are justified only if we stand in the right epistemic position with respect to them” (158). What does Lewis mean here by the right “epistemic position?”

Lewis argues that there are two “limiting conditions” on the appropriateness of blame, one Epistemic and one “Positional:” “we are justified in blaming others for their actions only to the extent that we have *evidence* that they acted on a morally objectionable attitude, and only to the extent that we are *in a moral position* to hold them to a standard that attitude fails to meet” (161, my emphasis). What we need, for blame to be appropriate, is to have good evidence of blameworthiness, and to have the right kind of standing to act on that evidence by blaming. The rest of Lewis’ argument attempts to show that we do not meet these limiting conditions as often as we think.

The basic insight driving the paper is that those who commit crimes are often incentivized to do so by the conditions in which they find themselves. They either do so because there are strong payoffs (in terms of whatever goods they find valuable), or because they think that there is a high likelihood of living a more overall valuable life if they do so. Importantly, this is *comparatively* true – their incentives are stronger than the advantaged, whatever the ultimate strength of the incentives is. This is important because it blocks an initial objection that *everyone* has some reason to engage in blameworthy behavior for illicit gain. This might be true, but if my incentives to do so are much weaker than social disincentives against committing crimes, then it is obvious that I am in a different kind of position from someone whose incentives are comparatively much stronger than those social disincentives.

This comparative claim is especially important in understanding the work that incentives do. That is, an easy objection to Lewis’ account says that, quite obviously, committing a crime which involves taking one’s own interests (however strong) to be more important than the comparable interests of victims is all the evidence of a bad will we’ll ever need. But, the epistemic limiting condition shows us that, in fact, committing a crime is not good evidence of a bad will when that crime is highly incentivized. At the very least, incentivization makes our evidence for a bad will *comparatively* weaker.

The positional move in Lewis’ picture is to note that “we” are often partially responsible for the incentive structure that low-status individuals find themselves in. Given its intended audience of academic philosophers, the paper’s inclusive use of “we” is probably justified here. But we can soften the claim and bring it into alignment with my own: those who wield a great deal of social power are partially responsible for the construction and maintenance of the very incentive structures that incentivize low-status individuals to commit crimes. Given this, their blame is (at least partially) inappropriate.

### 3 Responding to Objections

I’ve now argued that we aren’t very good at forming responsibility judgements across gaps in social power and position, that this is due to general epistemic difficulties, and that our practices will, therefore, be reliably faulty. Furthermore, I pointed out that our practices widely involve such gaps, and, therefore, ought to be revised. Finally, I’ve argued that the revision cannot focus *merely* on epistemic matters but must be sensitive to moral and pragmatic questions as well. The revision I’ve begun to describe involves introducing new

norms to our practice, and the particular norm I've outlined asks the powerful to refrain from blaming those with less social status across large gaps in power.<sup>26</sup>

This argument led to the principle of **Powerful Restraint** I introduced above. We ought to keep front of mind that the principle is not meant to depend on whether the powerful are likely to give up blaming the less powerful on their own - it is describing a normative aim. The idea here is to come up with a revisionary principle which can reorient and put normative pressure on everyone involved in our practices. The work of shifting norms, however, is incredibly complex and laborious - I don't mean to sell it short.

Let me try to make more precise the content and spirit of Powerful Restraint by dealing with a few further objections to it. To my mind, two main classes must be dealt with which I'll call Disrespect and Asymmetry.

**Disrespect:** Declining to blame those of lower social status is straightforwardly disrespectful. It denies them full membership in the moral community - treating them either with the Strawsonian objective stance, or as akin to children.

I think this objection is persuasive, but that it can be met. There is very interesting recent work on the connection between respect and responsibility - work that I find compelling.<sup>27</sup> One central idea is that by choosing to withhold blame, we are denying agents a certain kind of respect. Consider a non-moral case: a well-meaning teacher has a student with a learning disability in their class. Instead of providing the student accommodations that would make the classroom equitable, they simply grade the student less critically - declining to hold them accountable for their errors.

This is disrespectful, whether the teacher meant it to be or not. By denying the student an opportunity to be held accountable for their mistakes (in an environment in which it would have been fair to do so), the teacher is denying them full membership in the academic community, as well as the ability to improve their capacities.

So too in moral cases. By declining to blame those we might view as "less capable," we may be denying them access to our moral community, as well as the respect that goes with it, and the capacity to improve themselves. At a first pass, I think my view has a novel response to this kind of worry: In cases of Powerful Restraint, we are not declining to hold responsible because of a worry about the *agency* or *capacities* of the less powerful. Instead, we are declining to hold them responsible because of a worry about our *own* capacities.

It is harder to see how the charge of disrespect can stick in this situation. Imagine an analogy: if I am a surgeon, and I decline to perform a risky operation on you because I'm worried that I don't know enough about your symptoms to proceed, it is hard to see how this qualifies as disrespectful. Of course, we can imagine edge cases where someone of low-status *demand*s that we hold them responsible (or, in the analogy, someone demands that we proceed with the surgery) but this just looks like a case where we have sufficient evidence to override the generality clause in Powerful Restraint - there will still be cases where the evidence is good enough to blame.

Although it might be generally impermissible to blame across a large gap in power, it is no problem for the view if there are edge cases where permission *is* granted. I suppose the relevant question is how likely such edge cases are. There is no way to answer this *a*

<sup>26</sup> I've also hinted at the fact that the less powerful *ought* to continue to hold the powerful to account, and, indeed, that a second norm we may want to introduce into our practices is that they ought to do so more often, although I do not defend that norm here.

<sup>27</sup> For instance in Kathleen Connelly's (2021) "Blame and Patronizing," Manuscript submitted for publication, and Tommie Shelby's (2018) *Dark Ghettos*, Belknap Press of Harvard University Press, Cambridge -- particularly Part III

*priori*, but I can imagine one important set of cases that we should pay attention to. Return to our surgery analogy. One common complaint against our currently constituted practices of medicine in the United States is that they often discount the pain of women – and even more so the pain of women of color – particularly black women. We can imagine a case where a black woman is *sure* that she needs an elective surgery to reduce her pain, but a doctor refuses because of a lack of certainty about whether the procedure is necessary. This class of cases *may* involve the kind of disrespect that the objection was after.

So too in the realm of responsibility. If a low-status community commonly *complains* that those in power fail to respect their agency by holding them responsible, I would take it that this is good evidence that continuing to eschew accountability practices *would* be disrespectful. However, I am not aware of this kind of claim being brought forward with much frequency. Instead, the opposite claim, that the powerful are too quick to hold responsible, blame, and punish the less powerful is common, and precisely the issue I am endeavoring to deal with. So, as long as we are not in the class of cases where communities themselves are demanding to be held responsible, I think the charge of disrespect does not go through.

Before moving on, we should consider a closely related objection: that denying the less powerful the opportunity to be blamed robs them of opportunities for self development.<sup>28</sup> After all, on many plausible models of agency development, part of what helps us become competent practitioners of moral responsibility is our being “in the game,” so to speak.<sup>29</sup> We come to understand the relevant norms by coming into contact with them – we come to be competent blamers in part by learning when *we* are to blame. Not only this, blame can serve a valuable social function as a signal that wrongdoing has occurred, and, perhaps, as a form of moral protest.<sup>30</sup>

We can say (at least) two things in defense of Powerful Restraint against this kind of complaint. First, the low-status are not robbed of opportunities to be blamed full-stop. It is still the case that those of similar (and lower) social positions can and should blame them for blameworthy behavior. It is also still the case that, given that Powerful Restraint will be imperfectly followed, blame may flow down from above. So, it is not as if the low-status will suddenly live in a blameless world. A paper more focused on articulating a specific account of blame itself might also rely here on issues of standing. That is, I find it unlikely that, in many cases, the powerful will be in the best position to have unequivocal standing to blame. I won't argue this point at length here, but suffice it to say that I find it highly plausible that there are likely to be other members of a community who would be in better positions to blame in most cases. There will rarely, in other words, be overriding reasons for the powerful to step in in such cases, given the dangers I've canvassed. And again, if we can imagine a context where the powerful are the only ones able to engage in corrective or agency-enhancing blaming, and where such blame would truly be importantly agency-enhancing, the “in general” clause allows that such blame can be appropriate. It should be obvious, however, that I find it unlikely that this kind of situation will be common in our practices.

Second, and very briefly, there is also a definitional question of what counts as blame as opposed to nearby forms of corrective or enhancing critique. I have said that I wish to remain ecumenical about the nature of blame, and so I won't have much to say on this

<sup>28</sup> Thanks to an anonymous reviewer for pressing this point.

<sup>29</sup> For distinct and persuasive models of responsibility as a system of agency cultivation see: McGeer (2012), McGeer (2019), and Vargas (2013).

<sup>30</sup> See, for example, Hieronymi (2001).

matter. On some capacious definitions of blame, calm, dispassionate moral critique will count. It seems obvious that this kind of blame will be less likely to harm than full-throated emotional blame.<sup>31</sup> However, I merely point out here that on other accounts of blame, well-meaning (or even friendly) critiques will *not* count as paradigmatic of blaming. It is open, therefore, that some of the kinds of moral critique this objection imagines are still perfectly open to high-status individuals, downstream of some conceptual fights about what does or doesn't count as blaming. In all cases, however, we can ask: is it really the case that high-status individuals will be in the best position or have the best standing to blame?

The final major objection to be dealt with targets the asymmetry in blame's acceptability that I've introduced:

**Asymmetry:** Aren't the less powerful just as likely to err as the more powerful when it comes to epistemic processes concerning responsibility? And, going further, aren't there likely to be pernicious reasons unique to this context? If the poor want to send the innocent rich to the guillotine merely for existing, does my theory excuse this?

No answer I can give in the brief remainder of this paper will be fully satisfying.<sup>32</sup> Here's a sketch of how a longer answer would go: first we'd want to hear more about the likely outcomes of run-away responsibility ascription from low to high status. How likely is it that the rich are really going to the guillotine? Is a more likely outcome the re-distribution of wealth, or loss of opportunities for rich heirs? In any case, it looks like the verdict here is going to depend on instrumental calculations that are outside of my theory. Indeed, much here may hang on the supposed "innocence" of our metaphorical rich man. How likely is it that the billionaire is really blameless for society's ills? And, even if they *are* to blame for some of those ills, will the low-status appropriately constrain their blaming to the causes of those ills themselves, or be likely to condemn the billionaire more generally? Normative and ethical theories are going to guide us here as much as a theory of the epistemology of responsibility. My claim is merely that, given the lack of *power* possessed by the low-power in the first place, erring on the side of leniency is unobjectionable. The powerless should be free to form responsibility judgements and pursue their downstream effects precisely because: a) redistributing power (at least in non-violent ways) is not objectionable in these cases, and b) lacking power to begin with, such judgements are relatively unlikely to cause significant harm.

Second, we can, up to a point, engage in some bullet biting and say: instrumentalist justifications just price in certain kinds of errors. That is, the very point of the instrumental justification is that the practice is overall justified when set up in a certain way, while recognizing that there *will* be cases of error in the system. Combined with the first line of response, we'd say that holding the powerful accountable is unlikely to produce systematically bad results, even if it is occasionally done in error.

A related and equally thorny issue concerns *intra*-group blame. One common kind of claim is that those who are members of the same class or group of people are often their

<sup>31</sup> On the other hand, see "Blame Italian Style," Wolf (2011) for a defense of this kind of blame.

<sup>32</sup> And this is true along several dimensions. One thing that should be clear at this point in the paper is that fully working out the norms of who ought to constrain their blame (and when, and how much) would require a *much* more detailed working out of the relevant notion of social power we are working with. I've indicated at various points that this question is alive in the background of the paper. Will, for instance, the middle-class (as a fuzzy group) have sufficient power that individual members of that group should refrain from blaming the very poor in many instances? These are difficult questions that deserve further careful treatment, but I leave them aside in this paper.

own harshest critics. Given this, shouldn't we recommend that those within social groups blame each other less often? First of all, it is hard to know exactly how true such claims are - indeed, sometimes they seem motivated by sexism or other forms of bigotry. Consider the idea, for instance, that women are "catty" and mean to one another by nature, a claim which is surely false (or at the very least involves a sexist reading of a complex social schema). However, there are two reasons to think that *sometimes* intra-group members really are their own harshest critics - but both reasons, I think, militate against including them in a norm that restrains blame.

The first reason is simply the idea that one knows one's own group best, and is thus often in the best position to criticise it. This is at least sometimes true. However, if this is the reason that group members hold each other accountable more often or more harshly, this would be so because such judgements are *accurate*. In such cases, then, members would *not* be in error, and so there would be no reason for them to restrain their blame *based on my view*.

The second reason is that these intra-group judgements of responsibility may involve a sublimation of the very norms of oppression that the powerful wield. In other words, group members may be quick to blame each other because of internalized misogyny, racism, or cases of adaptive preference. This *would* involve moral and epistemic error, and mean that there *is* a reason for restraint in these kinds of cases.

But, notice two things: first, it is unlikely that those who are acting out of internalized misogyny or adaptive preferences will notice what they are doing, or, insofar as they do, describe it in those terms. So a principle that asks them to change their behavior is unlikely to be effective (and this is putting aside other thorny issues about asking those with adaptive preferences to change them).

Second, the aim of Powerful Restraint is to change the normative landscape. If what we are witnessing is really a "trickle down" moral universe where the oppressed are taking up the norms of their oppressors, then changing the norms at the top will (eventually) put an end to it. I am not, of course, claiming that anything like this is easy. The idea is just that, as more social pressure is put on the powerful to restrain their blame, there may be opportunities for those who are internalizing oppression to come to see what is occurring to them *as* oppressors.

In either case, then, I argue that Powerful Restraint would be a more effective tool than trying to constrain intra-group blame. However, I think it's fair to say that these last few objections point us back towards the much more general upshot that a careful working out of the epistemology of responsibility raises: figuring out how to accurately form judgements about responsibility is hard – harder, at least, than we may have first believed.

## Declarations

**Conflict of Interests** The author has no relevant financial or non-financial interests to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abizadeh, Aresh (2021). "The Grammar of Social Power". In: *Political Studies*, pp. 1–17. ISSN: 00016993
- Alicke MD, Buckingham J et al (2008) Culpable control and counterfactual reasoning in the psychology of blame. *Personal Soc Psychol Bull* 34(10):1371–1381. ISSN: 01461672. <https://doi.org/10.1177/0146167208321594>
- Alicke MD, Mandel DR et al (2015) Causal conceptions in social Explanation and moral evaluation: a historical tour. *Perspect Psychol Sci* 10(6):790–812. ISSN: 17456924. <https://doi.org/10.1177/1745691615601888>
- Barden J et al (2004) Contextual moderation of racial bias: the impact of social roles on controlled and automatically activated attitudes. *J Pers Soc Psychol* 87(1):5–22. ISSN: 00223514. <https://doi.org/10.1037/0022-3514.87.1.5>
- Bayles MD (1982) Character, purpose, and criminal Responsibility. *Law Philos* 1:5–20
- Brewer MB (1977) An information-processing approach to attribution of responsibility. *J Exp Soc Psychol* 13(1):58–69. ISSN: 10960465. [https://doi.org/10.1016/0022-1031\(77\)90013-0](https://doi.org/10.1016/0022-1031(77)90013-0)
- Brink, David O and Dana Kay Nelkin (2013). "Fairness and the architecture of Responsibility". In: *Oxford Studies in Agency and Responsibility*, pp. 284–314
- Buunk BP, Mussweiler T (2001) New directions in social comparison research. *Eur J Soc Psychol* 31(5):467–475. ISSN: 00462772. <https://doi.org/10.1002/ejsp.77>
- Calhoun, Cheshire (1989). "Responsibility and Reproach". In: *Ethics* 99.2, pp. 389–406.
- Cudd, Ann E. (2006). "Analyzing Oppression". In: *Analyzing oppression*, pp. 1–304. <https://doi.org/10.1093/0195187431.001.0001>
- Cuddy AJC, Fiske ST, Glick P (2007) The BIAS map: behaviors from intergroup affect and stereotypes. *J Pers Soc Psychol* 92(4):631–648. ISSN: 00223514. <https://doi.org/10.1037/0022-3514.92.4.631>
- Devine, D.J. and D.E. Caughlin (2014). "Do they matter? A meta-analytic investigation of individual characteristics and guilt judgments". In: *Psychology and Public Policy Law* 20.109
- Dotson, Kristie (2014). "Conceptualizing Epistemic Oppression". In: *Social Epistemology* 28.2, pp. 115–138. ISSN: 14645297. <https://doi.org/10.1080/02691728.2013.782585>. <https://doi.org/10.1080/02691728.2013.782585>
- Ellison L, Munro V (2008) Reacting to rape: exploring mock jurors' assessments of complainant credibility. *British Journal of Criminal Law* 49:202–219
- Feigenson N (2016) Jurors' emotions and judgments of legal Responsibility and blame: what does the experimental research tell us? *Emot Rev* 8(1):26–31. ISSN: 17540739. <https://doi.org/10.1177/1754073915601223>
- Fischer JM, Ravizza M (1998) Responsibility and control: a theory of moral Responsibility. Cambridge University Press, Cambridge
- Fiske ST, Cuddy AJC, Glick P (2007) Universal dimensions of social cognition: warmth and competence. *Trends Cogn Sci* 11(2):77–83. ISSN: 13646613. <https://doi.org/10.1016/j.tics.2006.11.005>
- Fiske ST, Cuddy AJC, Glick P, Xu J (2002) A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *J Pers Soc Psychol* 82(6):878–902. ISSN: 00223514. <https://doi.org/10.1037/0022-3514.82.6.878>
- Fricker M (2007) Epistemic injustice: power & the Ethics of knowing. Oxford University Press, Oxford
- Gailey JA, Frank Falk R (2008) Attribution of responsibility as a multidimensional concept. *Social Spectr* 28(6):659–680. ISSN: 15210707. <https://doi.org/10.1080/02732170802342958>
- Gawronski B (2009) The multiple inference model of social perception: two conceptual problems and some thoughts on how to resolve them. *Psychol Inq* 20(1):24–29. ISSN: 1047840X. <https://doi.org/10.1080/10478400902744261>
- Gerber JP, Wheeler L, Suls J (2018) A social comparison theory meta-analysis 60+ years on. *Psychol Bull* 144(2):177–197. ISSN: 00332909. <https://doi.org/10.1037/bul0000127>
- Hieronimi P (2001) Articulating an Uncompromising Forgiveness. In: *Philosophy and Phenomenological Research* 62.3, pp. 529–555. <https://doi.org/10.2307/2653535http://www.jstor.org/stable/2653535%5Cnhttp://www.jstor.org/stable/pdfplus/2653535.pdf>
- Krueger J (2000) The projective perception of the social world. In: Suls J, Wheeler L (eds) *Handbook of social comparison*. Kluwer Academic Publishers, New York
- Lagnado DA, Channon S (2008) Judgments of cause and blame: the effects of intentionality and foreseeability. *Cognition* 108(3):754–770. ISSN: 00100277. <https://doi.org/10.1016/j.cognition.2008.06.009>
- Lewis C (2016) Inequality, incentives, criminality, and blame. *Legal Theory* 22(2):153–180. ISSN: 14698048. <https://doi.org/10.1017/S1352325217000052>
- Manne K (2018) Down girl: the logic of misogyny. Oxford University Press, Oxford ISBN: 978-0-19-060498-1



- Mazella R, Feingold A (1994) The effects of physical attractiveness, race, socioeconomic status, and gender of defendants and victims of judgments of mock jurors: a meta analysis. *J Appl Soc Psychol* 24:1315–1338
- McGeer V (2012) Co-reactive attitudes and the making of moral community. *Emotions, Imagination, and Moral Reasoning* 1974:299–326. <https://doi.org/10.4324/9780203803134>
- McGeer V (2019) Scaffolding agency : a proleptic account of the reactive attitudes. *European Journal of Philosophy* 2018:301–323. <https://doi.org/10.1111/ejop.12408>
- McKenna M (2012) *Conversation and Responsibility*. Oxford University Press, New York
- Menge T (2020) Fictional expectations and the ontology of power. In: *Philosophers imprint* 20.29, pp. 1–22. ISSN: 1533628X
- Mills CW (2017) *Black rights/ white wrongs*. Oxford University Press, Oxford
- Mitchell TL et al (2005) Racial bias in mock juror decision-making: a meta-analytic review of defendant treatment. *Law Hum Behav* 29:621–637
- Nadler J (2012) Blaming as a social process: The influence of character and moral emotion on blame. In: *Law and Contemporary Problems* 75.2, pp. 1–31. ISSN: 00239186. [https://www.law.northwestern.edu/faculty/fulltime/nadler/Nadler\\_LCP\\_2012.pdf](https://www.law.northwestern.edu/faculty/fulltime/nadler/Nadler_LCP_2012.pdf)
- Nadler J, McDonnell MH (2012) Moral character, motive, and the psychology of blame. In: *Cornell Law Review* 97.2, pp. 255–304. ISSN: 00108847
- Oshana M (2018) Ascriptions of Responsibility given commonplace relations of power. In: *social dimensions of moral Responsibility*. Ed. by Katrina Hutchison, Catriona Mackenzie, and Marina Oshana. Oxford: Oxford University press. Chap. 3, pp. 81–110
- Pickard H (2013) Responsibility without blame: philosophical reflections on clinical practice. In: *Cure and Care* 2010, pp. 1134–1154
- Pizarro DA, Tannenbaum D (2012) Bringing Character Back: How the Motivation to Evaluate Character Influences Judgments of Moral Blame. In: *The Social Psychology or Morality: Exploring the causes of good and evil*. Ed. by Mario Mikulincer and Philip R. Shaver. Washington, DC: American Psychological Association. Chap. 5, pp. 91–108. ISBN: 9788578110796. [10.1017/CBO9781107415324.004](https://doi.org/10.1017/CBO9781107415324.004). arXiv: arXiv:1011.1669v3
- Rahimi S, Hall NC, Pychyl TA (2016) Attributions of responsibility and blame for procrastination behavior. In: *Frontiers in Psychology* 7.AUG, pp. 1–7. ISSN: 16641078. <https://doi.org/10.3389/fpsyg.2016.01179>
- Scanlon TM (2008) *Moral Dimensions*. Cambridge, MA: Belknap Press of Harvard University Press. ISBN: 9788578110796. <https://doi.org/10.1017/CBO9781107415324.004>. arXiv: arXiv:1011.1669v3
- Shaver KG (1985) The attribution of blame: causality, responsibility, and blameworthiness. New York: Springer. ISBN: 9781461295617 <https://doi.org/10.1007/978-1-4612-5094-4>
- Smith A (1976) *The theory of moral sentiments*. Ed. by A.L. MacFie and D.D. Raphael. The Glasgow. Oxford: Oxford University Press. ISBN: 9781626239777
- Sommers SR, Ellsworth PC (2000) Race in the courtroom: perceptions of guilt and dispositional attributions. *Pers Social Psychol Bull* 26:1367–1379
- Strawson P (2003) Freedom and resentment. In: *Free Will* Ed by Gary Watson Oxford: Oxford University Press Chap 4, pp. 72–93
- Suls J, Martin R, Wheeler L (2002) Social comparison: why, with whom, and with what effect? *Curr Dir Psychol Sci* 11(5):159–163. ISSN: 09637214. <https://doi.org/10.1111/1467-8721.00191>
- Suls J, Wheeler L (2000) A selective history of classic and neo-social comparison theory. In: *Handbook of Social Comparison*, pp. 3–19. [10.1007/978-1-4615-4237-7\\_1](https://doi.org/10.1007/978-1-4615-4237-7_1)
- Vargas MR (2013) *Building better beings: a theory of moral responsibility*. Oxford University Press
- Westra E (2018a) Character and theory of mind: an integrative approach. *Philos Stud* 175(5):1217–1241. ISSN: 15730883. <https://doi.org/10.1007/s11098-017-0908-3>
- Westra E (2018b) Character and theory of mind: an integrative approach. *Philos Stud* 175(5):1217–1241. ISSN: 15730883. <https://doi.org/10.1007/s11098-017-0908-3>
- Westra E (2019) Stereotypes, theory of mind, and the action–prediction hierarchy. *Synthese* 196(7):2821–2846. ISSN: 15730964. <https://doi.org/10.1007/s11229-017-1575-9>
- Willemsen P, Newen A, Kaspar K (2018) A new look at the attribution of moral responsibility: the underestimated relevance of social roles. In: *Philosophical Psychology* 31.4, pp. 595–608. ISSN: 1465394X. <https://doi.org/10.1080/09515089.2018.1429592>
- Wolf, Susan (2011). “Blame, Italian style”. In: *Reason and recognition: essays on the philosophy of T.M. Scanlon*. Ed. by R. Jay Wallace, Rahul Kumar, and Samuel Freeman. New York: Oxford University Press, pp. 332–347
- Young IM (1990) *Justice and the politics of difference*. Princeton: Princeton University Press. ISBN: 0691078327. <https://doi.org/10.1177/1522637916656379>
- Zhao J, Rogalin CL (2017) Heinous crime or unfortunate incident: does gender matter? *Soc Psychol Q* 80(4):330–341. ISSN: 01902725. <https://doi.org/10.1177/0190272517728923>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.