Episteme (2025), 1–20
doi:10.1017/epi.2025.10063

episteme

ARTICLE

Trust Reductions, Epistemic Blame, and Preventative Measures

Roman Heil

Goethe University Frankfurt, Frankfurt, Germany
Email: r.heil@em.uni-frankfurt.de

(Received 9 February 2025; revised 12 June 2025; accepted 26 June 2025)

Abstract

When we learn that someone holds irrational beliefs, we often respond by reducing our epistemic trust in them. In this paper, I will propose a novel account of such trust reductions. The recently popular relationship-modification account (RMA) of epistemic blame will serve as a foil for this project. RMA says that epistemically blaming others for their epistemic failings involves modifying our epistemic relationships with them, paradigmatically via a reduction of epistemic trust. RMA has recently faced two challenges of extensional inadequacy, which I will show result from a mistaken view about what type of response trust reductions are. I will draw on resources from legal theory to show that trust reductions bear all the hallmarks of so-called non-punitive measures, which serve preventative collective purposes. I will argue for an account of trust reductions as *non-punitive epistemic measures* that serve the purpose of preventing unreliable informants from negatively affecting the epistemic commons. My account explains when and why it is appropriate to reduce epistemic trust and shows where RMA goes wrong.

Keywords: Epistemic trust; epistemic blame; relationship modification; epistemic accountability; epistemic commons; collectivist metaepistemology

1. Introduction

We often respond to epistemic failings by reducing epistemic trust. For illustration, consider the following case.

> **Fortune** Amari's colleague Stephen tells her about his investment strategy over lunch: whenever he considers buying stocks, he first consults his fortune-teller. Amari tells him that that is a terribly unreliable way of forming beliefs about financial matters, but Stephen remains unfazed. She decides not to trust his financial advice going forward.

In response to finding out that Stephen has formed – and will continue to form – beliefs about financial matters in an unreliable way, Amari reduces her trust in him concerning these matters. We respond similarly to those who epistemically fail in other ways, for

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/epi.2025.10063 Published online by Cambridge University Press

example, by holding beliefs based on wishful thinking or dogmatism, or by engaging in sloppy or motivated reasoning.

This paper proposes a novel account of epistemic trust reductions.[1] Let's start out with an initial characterization of the phenomenon I am interested in (cf. Kauppinen 2018; Boult 2024c; in what follows, I will drop the 'epistemic' for the sake of brevity). First, trust reductions are responses to epistemic failings, that is, to negatively evaluated epistemic conduct. Second, trust reductions are typically restricted to some domain or topic. Suppose that a famous physicist believes, based on motivated reasoning, that medical face masks have no protective effects against airborne illnesses. While we should reduce our trust in him when it comes to medical topics, it might still be appropriate to not reduce trust in his domain of expertise in physics, at least absent further evidence that there, too, he engages in bad kinds of reasoning.[2] Third, a reduction of trust can manifest in various ways. Typically, we are less inclined to believe what the target subject says, less willing to partner with them in collaborative inquiry, and less confident in their abilities to form true beliefs on the relevant topic in the future, and so on.

Many have recently argued that trust reductions are a distinctively epistemic kind of response to epistemic failings. For instance, Antti Kauppinen (2018, 2023) has argued that reducing trust is a way for us to hold each other accountable for violating epistemic norms. According to Cameron Boult's recently popular relationship-modification account (RMA) of epistemic blame, trust reductions are paradigmatic ways of epistemically blaming others and ourselves for epistemic failings (Boult 2020, 2023, 2021a, 2021b, 2024b, 2024c, 2024d; Schmidt 2024; Flores and Woodard 2023; Woodard 2024).

RMA and the central role it assigns to trust reductions will serve as a foil for my project of proposing a novel account of trust reductions. RMA has faced two challenges of extensional inadequacy, which I will argue both ultimately arise because RMA incorrectly conceptualizes trust reductions as paradigmatic instances of epistemic blame. First, it has been argued that RMA counterintuitively counts as epistemic blame certain kinds of positively valenced responses, such as showing curiosity or understanding (cf. Smith 2013; Chislenko 2020; Boult 2024c, §3.2). Second, it has been argued that RMA counterintuitively counts as epistemic blame reductions of trust in non-culpable, and thus intuitively not blameworthy, agents (Smartt 2023; Piovarchy forthcoming). I will draw on resources from legal theory to argue that trust reductions are not instances of epistemic blame, but rather bear all the hallmarks of so-called non-punitive measures, that is, types of sanctions that exclusively serve preventative and (sometimes) rehabilitative purposes. I will develop a novel, explanatorily powerful account of trust reductions as *non-punitive epistemic measures* (NoPEMs) and then put it to work to assess RMA's challenges and extant responses. While my account motivates principled responses to the two challenges, it also provides another, more general reason for being skeptical of RMA.

The paper is structured as follows. In section 1, I will present RMA with its various moving parts. In section 2, I will present the two challenges to RMA. In section 3, I will draw on legal theory and legal codes to introduce the notion of a non-punitive measure and elaborate on the latter's application conditions. In section 4, I will propose a novel account of epistemic trust reductions which says that they should be conceptualized as NoPEMs. In section 5, I will use my account to assess RMA's two challenges and extant

[1]I will only be concerned with *epistemic* trust, which plausibly differs from practical trust (see Kauppinen 2018, p. 6).

[2]For possible real-life examples, see Ballantyne (2019).

https://doi.org/10.1017/epi.2025.10063 Published online by Cambridge University Press

responses. In section 6, I will take stock of the options remaining for RMA and consider some alternatives. Section 7 concludes.

2. The relationship-modification account of epistemic blame

In what follows, I will present the RMA of epistemic blame and elaborate on the central role it assigns to trust reductions. Let's start with an initial characterization of epistemic blame. It is commonly regarded as a response to an epistemic failing, not to its practical downstream effects (Boult 2024c, p. 31). Epistemic blame is taken to go beyond mere negative evaluation in that it involves a particular 'sting, force or depth' (Hieronymi 2004, pp. 116–7) and in that the blamer is *engaged* by his response (Piovarchy 2021, p. 794; Boult 2020, p. 519; 2024c, 19). Finally, some have argued that epistemic blame typically does not involve 'hot' reactive attitudes (e.g., indignation or resentment), but rather 'cool' responses (Boult 2020), such as trust reductions.

The RMA of epistemic blame has been extensively developed by Cameron Boult in his recent book and a series of papers. RMA is inspired by Scanlon's (2008) relationship-modification account of *moral* blame[3] and has various moving parts, which I will present in what follows.[4] Here is a first approximation of RMA's conception of epistemic blame (Boult 2024c, p. 66).[5]

> **Epistemic Blame** One's response to an agent *S*'s epistemic failing constitutes epistemic blame iff one modifies one's epistemic relationship with *S* in a way that fits one's judgment that *S* is epistemically blameworthy.

Let's unpack. According to RMA, first, one modifies one's epistemic relationship to *S* by modifying one's intentions and expectations toward *S*. The paradigmatic kind of modification is the *reduction of epistemic trust* (Boult 2020, p. 521, 2021a, p. 6, 2023, p. 817, 2024b, p. 391; Woodard 2024, p. 172; Schmidt 2024, p. 18; Flores and Woodard 2023, p. 2557). For instance, if we epistemically blame *S* by reducing trust in her regarding some topic *t*, we expect her to be an unreliable informant with regard to *t* and intend to lend less credence to her *t*-regarding assertions. Second, a judgment that *S* is epistemically blameworthy is understood to be a judgment that *S* has done something that impairs their epistemic relationship with the judging agent. Third, *S* has impaired an epistemic relationship iff S falls short of the normative ideal of that relationship. Kicking the conceptual can just one step further down the road, the normative ideal of epistemic relationships is thought to be constituted by both general and particular epistemic expectations.[6] Drawing on work from Goldberg (2017, 2018), Boult assumes that we have a general epistemic expectation that others are both epistemically trustworthy and trusting, that is, that they are reliable informants and defer to other reliable informants under appropriate circumstances. There is a corresponding practice of mutual epistemic

[3]It is not straightforward to nail down Scanlon's view, but neither Boult (2024c, p. 102) nor I in what follows are concerned with Scanlon exegesis. For an excellent overview over the interpretive options, see Chislenko (2020).

[4]Besides RMA, there are emotion-based accounts of epistemic blame (Nottelmann 2007; McHugh 2012; Rettler 2018), desire-based accounts (Brown 2018, 2020), and agency-cultivating accounts (Piovarchy 2021). See Boult (2021a) for an overview.

[5]Boult has modified his account at various points; I will stick to the newest presentation of RMA from his recent book (2024c). I will consider some of Boult's refinements to Epistemic Blame later on.

[6]These expectations are both predictive and normative, constituting an entitlement (Goldberg 2018, ch. 5) that there is a mutual adherence to these ideals. See Brown (2020) for the worry that these general expectations are only predictive, not normative, and Boult (2024c, pp. 151–4) for a response.

https://doi.org/10.1017/epi.2025.10063 Published online by Cambridge University Press

trust in each other's testimony, the purpose of which is to facilitate the division of epistemic labor and protect the epistemic commons. Additionally, there are also more particular epistemic expectations – between, for example, teachers and students, doctors and patients, politicians and voters, friends, family, and so on (Boult 2023, p. 820, 2024c, ch. 6.5) – that are grounded in professional, institutional, or miscellaneous other relationships.

For illustration, let's apply RMA to our initial case, Fortune. Stephen falls short of the normative ideal of his epistemic relationship with Amari by consulting and believing the fortune-teller concerning financial matters. He violates Amari's (and general) epistemic expectations by using an unreliable method to form beliefs about the topic in question, namely by trusting the testimony of a fortune-teller. Amari judges that Stephen has thereby impaired his epistemic relationship with her, and she fittingly adjusts her intentions (to not believe him on financial matters) and her expectations toward him (that he will be unreliable on financial matters in the future) to reflect this, thereby reducing epistemic trust. In doing so, according to RMA, Amari epistemically blames Stephen for his epistemic failing.

Epistemic relationship modifications seem to be well-suited to capture the abovementioned characteristics of epistemic blame. First, the modification is made fitting by one's judgment about the epistemic failing, not its downstream effects.

Second, RMA explains what separates epistemic blame from mere negative evaluation: holding others to normative expectations and altering one's epistemic relationship if these expectations are frustrated goes beyond giving someone 'a bad epistemic grade' (Boult 2024c, 83); rather, it is a way of being engaged by one's response to epistemic failings (ibid., §4.1). Finally, RMA explains the putatively 'cool' nature of epistemic blame: as Fortune and other cases illustrate, 'cool' -headed trust reductions are often a fitting way to modify one's epistemic relationships.

RMA's account of epistemic blame has faced two major challenges from extensional inadequacy. I will briefly introduce these challenges now and consider extant responses in more detail later on. The first challenge is the epistemic analog of a well-known challenge from Smith (2013, p. 375) to Scanlon's account. As Chislenko (2020, p. 375) puts the original worry, the relationship-modification account says that

> any modification of a relationship can count as blame, as long as the judgment of blameworthiness holds it to be appropriate. Reacting to [our friend] Joe[, who made a cruel joke about us,] with love, emotional support, or curiosity may not, on this view, be alternatives to blame. They could *be* blame. This seems to stretch our conception of blame beyond borderline cases, and, indeed, beyond recognition.

Likewise, if we respond in similarly positively valenced ways to epistemic failings, for example, with curiosity or understanding, and judge these responses to be fitting (i.e., judge that they are fitting responses to someone's epistemic φ-ing falling short of the normative ideal of the respective epistemic relationship), then RMA would count them instances of epistemic blame. In Fortune, for instance, it seems fitting to respond to Stephen with curiosity: why does he believe a fortune-teller, and why only concerning financial matters? Likewise, Boult (2024c, p. 185) discusses a case in which one is concerned about a friend who has been led down a YouTube conspiracy rabbit hole, and where showing understanding can be a fitting response to the friend's epistemic failing (although Boult argues that this response doesn't constitute epistemic blame; more on this later). But responding with curiosity or understanding to someone's epistemic failing is intuitively not a way of epistemically blaming them.

https://doi.org/10.1017/epi.2025.10063 Published online by Cambridge University Press

The second challenge has been forcefully raised by Smartt (2023, p. 1820f) and concerns trust reductions in non-culpable agents (see also Piovarchy forthcoming). Consider the following case from Greco (2021, p. 531):

> **Senile Mother** 'Suppose I learn that my aging mother was taken in by an email scam—she sent a great deal of money to someone claiming to be a Nigerian prince who needed assistance transferring funds out of the country, and who promised to pay her back a hundredfold. Given the right background conditions—senility, mainly—I probably won't blame her for her belief that she was sending money to a genuine prince. But, knowing that she made a mistake like this, I will take her judgment less seriously going forward—when she says she's found a promising investment opportunity, I'll be less inclined to trust, and more inclined to verify'.

We judge that the mother has fallen short of general epistemic expectations and fittingly modify our epistemic relationship to her by reducing epistemic trust. Generally speaking, we sometimes seem to fittingly respond to the epistemic failings of non-culpable agents by reducing trust in them. Yet, it is implausible that we thereby blame them. Again, various responses have been given to this challenge. I will consider one of them now, and the others in more detail later on.

The response in question claims that we don't reduce epistemic trust in non-culpable agents. Both Schmidt (2024, §4.2) and Boult (2024b, fn.11) have suggested that we might merely be less inclined to *rely* on non-culpable agents, without reducing trust in them. This proposal strikes me as a nonstarter: the orthodox view is that trust is reliance 'plus some extra factor' (Hawley 2014, p. 5; Goldberg 2020).[7] Hence, there can't be a reduction in one's willingness to rely without a reduction of one's willingness to trust. Those who deny this owe us an alternative story about the relation between epistemic trust and reliance.

I will argue that ultimately, both challenges arise because RMA incorrectly conceptualizes trust reductions as paradigmatic instances of epistemic blame. According to RMA, trust reductions are paradigmatic ways of blaming ourselves and others for epistemic failings. In what follows, I will develop a novel account of trust reductions that shows that this view is mistaken.

Let me briefly sketch where I think RMA goes wrong. Proponents of RM think that trust reductions are paradigmatic epistemic blame responses because we often respond with trust reductions to culpable epistemic failings. However, I will argue that the latter is insufficient evidence for thinking that trust reductions (at least those fitting a corresponding judgment of blameworthiness) are paradigmatic manifestations of epistemic blame. Rather, by drawing on resources from legal theory, I will show that there is excellent evidence to think that trust reductions have been misidentified as (broadly conceived) culpability-tracking responses and rather should be understood as distinct responses – non-punitive measures – that can *accompany* culpability-tracking responses, but are not (partially) constitutive of the latter. My novel account of trust reductions as NoPEMs gets this right and offers a principled way to address RMA's challenges and assess extant responses.

[7]But see Thompson (2017) for possible cases of *practical* trust without reliance. As far as I can see, they are not readily translatable to the epistemic domain.

https://doi.org/10.1017/epi.2025.10063 Published online by Cambridge University Press

3. Non-punitive measures

In what follows, I will draw on rich conceptual resources from legal theory and legal codes to elaborate this dual system of responses. The mediate purpose of doing so is to extract characteristics and general application conditions for non-punitive measures. This prepares my subsequent argument that trust reductions exactly match these characteristics and application conditions in the epistemic domain.

Both common law traditions and (particularly) continental law traditions recognize, to various extents, a dual system of responses to legal 'failings', that is, unlawful conduct.[8] The first, familiar kind of *punitive sanctions* (fines, jail time, and so on) serve, depending on your favorite theory of punishment, the purposes of either retribution, restitution, denunciation, harm reduction (e.g., by deterrence), rehabilitation, the strengthening of social values or norm recognition, or, commonly, a pluralist mix of these.[9] Punitive sanctions are responses to culpable unlawful conduct. They impose on the culpable agent negative consequences that are ordinarily considered to be unpleasant (Dressler 2015, §2.02). They thereby express a negative evaluation of the agent and their conduct, where the strength of the negative evaluation is thought to correspond to the severity of punishment.

The second, perhaps less familiar, type of response is *non-punitive measures* ('non-punitive sanctions'), which are also responses to unlawful conduct that impose negative consequences. However, non-punitive measures don't presuppose culpability. They also don't express a negative evaluation of the agent and their conduct. The sole purpose of non-punitive measures is the individualized ('special') *prevention* of future unlawful conduct by the agent and (sometimes) also the agent's *rehabilitation*. For instance, consider a schizophrenic agent who suffers from severe hallucinations and assaults others because of his condition. Responding to his unlawful conduct by committing him to a psychiatric hospital not only aims at preventing further unlawful conduct caused by his condition, but also aims at therapeutically rehabilitating the agent. Importantly, it is sometimes appropriate to respond to unlawful conduct with *both* punitive sanctions and non-punitive measures. Being a response to unlawful conduct is hence insufficient to classify a response either way. Rather, we also need to look at the respective responses' features and purposes.

I will argue that epistemologists would benefit from taking notice of the rich and sophisticated dual classification of responses to unlawful conduct in legal theory and codes. Once we do, we will see that classifying trust reductions as epistemic blame responses is erroneous.

Taking inspiration from other normative domains, most notably ethics, to develop accounts of responses to epistemic failings is not a new approach (cf. Boult 2021a, §3.1). For instance, Brown (2018, 2020), Piovarchy (2021), and Boult (2020, 2023, 2021b, 2024b,c) have drawn on accounts of moral blame from Sher (2005), Vargas (2013), and Scanlon (2008), respectively, to develop accounts of epistemic blame. Why draw on legal theory? One reason is the noteworthy *social turn* in epistemology and the increasing interest in our epistemic practices. A common thought is that our epistemic assessments are governed by epistemic norms that are grounded in social practices, the purpose of which is to facilitate our division of epistemic labor and protect the epistemic

[8]For the perspective of US and German criminal law see respectively Dressler (2015, §2.02) and Meier (2019, §2.1.1, §5.1), and Dubber and Hörnle (2014, pp. 58–72) for a comparison.

[9]See Wood (2010a,b) for an overview.

https://doi.org/10.1017/epi.2025.10063 Published online by Cambridge University Press

commons.[10] In the light of this metaepistemological backdrop, the analogy to the legal domain is only natural. Legal conceptualizations of our responses to unlawful conduct are grounded and refined in long-running social practices. These practices aim, among other things, at facilitating and protecting social commons. Furthermore, legal practices are often explicitly codified to a high degree of systematicity and sophistication, making them particularly powerful conceptual blueprints to draw inspiration from in epistemological theorizing. These broadly Austinian (1957, p. 8) reasons suggest that epistemologists would benefit from taking an interest in legal conceptualizations and classifications of responses to failings.[11]

With the general approach motivated, let me now introduce the notion of a non-punitive measure in more detail. For illustration, I will draw on the particularly rich codification of non-punitive measures in the German criminal code (StGB).[12] Presenting them selectively in more detail will serve the aim of extracting their general characteristics and application conditions, which I will then deploy in the epistemic domain in the next section.

Consider again being committed to a psychiatric facility (§63), which is (very roughly) appropriate only if an agent committed an unlawful act, is not (fully) culpable for their conduct due to various psychiatric, drug-related (e.g., addictions) or other cognitive conditions (e.g., intellectual disabilities) and is expected to commit further (severe) crimes because of their condition. The specific purpose of this non-punitive measure is to prevent further unlawful conduct caused by the condition and, if feasible (van Gemmeren 2020), to improve the condition by therapeutic treatment.

While commitment to a psychiatric facility presupposes the absence of (full) culpability for the committed unlawful conduct, non-punitive measures do not generally presuppose this. Consider, for instance, the not uncontroversial non-punitive measure that is preventative detainment (§66). Depriving an agent of his liberty via preventative detainment is (very roughly) appropriate only if, among other things, the agent has already been sentenced for severe crimes in the past, is sentenced for another severe crime, and, based on an overall evaluation of his character and his crimes, is expected to be a continued danger to society.

Or consider the loss of one's driver's license (§69), which can be assigned irrespective of culpability and does not involve a rehabilitative dimension.[13] This non-punitive measure is appropriate for agents who are expected to not safely operate their vehicle because, for instance, they manifest certain excusing conditions (e.g., certain psychiatric or medical conditions) or display a lack of 'aptitude of character' (e.g., by engaging in illegal street racing).

While non-punitive measures differ in various ways from each other, some common, general application conditions can be extracted. First, non-punitive measures are

[10]This kind of social epistemic instrumentalism has most recently been endorsed by Dyke (2021), Chrisman (2022), Fleisher (2025) and Hannon and Woodard (forthcoming). For kindred views, see Williams (1973), Craig (1990), Neta (2006), Reynolds (2002), Fricker (2007), Greco (2007), Graham (2015), Henderson (2009, 2011), Rysiew (2012), Dogramaci (2012, 2015), Dogramaci and Horowitz (2016), Greco and Hedden (2016), McGrath (2015), Goldberg (2017, 2018), Hannon (2013, 2015, 2019) and Kusch and McKenna (2020).

[11]See Greco (2021) for pioneering work in this regard.

[12]The full list includes being committed to a psychiatric or rehabilitation facility (§§63–64), preventative detainment (§66), assignment of supervision of conduct (§68), loss of driver's license (§69), and occupational ban (§70). See Meier (2019, §5) for a detailed overview. Needless to say, my presentation will be extremely compressed and will gloss over many legal details that are orthogonal to my discussion.

[13]Even though there are mediate rehabilitative effects due to agents' typical interest in reacquiring their license (Meier 2019, p. 285).

https://doi.org/10.1017/epi.2025.10063 Published online by Cambridge University Press

responses to *unlawful*, that is, negatively evaluated, conduct that, second, impose negative consequences ordinarily considered unpleasant. Third, non-punitive measures, in general, can be appropriate irrespective of culpability, that is, they can be applicable in response to both culpable and non-culpable unlawful conduct. Importantly, they can *accompany* punitive sanctions. Suppose *A* engages in illegal street racing with their car and gets caught by the police. Not only might *A* get punitively sanctioned, but they might also be subject to a non-punitive measure: engaging in illegal street racing might reveal that they lack the needed awareness or care for the risk involved in driving, making it appropriate to non-punitively deprive them of their driver's license. Fourth, there needs to be a (reasonable) expectation of future unlawful conduct, which is grounded in an 'overall evaluation' of the agent. Indeed, the initial unlawful conduct has to *reveal* something about the agent's character, dispositions, and social environment in virtue of standing in a 'symptomatic relation' (Meier 2019, p. 299) to the expected future unlawful conduct.

Non-punitive measures are tightly constrained by considerations of *proportionality*. One such constraint is that the expected unlawful conduct has to be sufficiently likely and sufficiently severe. For instance, preventative detention is a grave interference with one's rights and liberties; so, it is only an appropriate response if the expected unlawful conduct surpasses particularly high barriers concerning likelihood and severity of the expected unlawful conduct.[14] Relatedly, proportionality requires that if the condition that caused or contributed to the unlawful behavior was only temporary (or ceases to persist at some point), then nonpunitive measures are not appropriate (or their continued appropriateness has to be reevaluated and the measure possibly suspended).

Finally, although the primary and characteristic purpose of non-punitive measures is prevention, some non-punitive measures have the subordinate purpose of rehabilitation. Whether this subordinate purpose translates into rehabilitative efforts, and if so, to what extent, depends on the particular measure and the likelihood of successful rehabilitation.[15]

Summing up, non-punitive measures are (1) responses to unlawful conduct, (2) impose negative consequences, (3) don't generally presuppose culpability, and (4) are only appropriate if there is a persistent expectation of sufficiently likely and severe future unlawful conduct. (5) They also sometimes involve a rehabilitative dimension.

In the next section, I will present a novel account of trust reductions, according to which trust reductions are NoPEMs. Besides showing that my account is explanatorily powerful, I will argue that, tellingly, the application conditions for trust reductions identified by proponents of RMA straightforwardly match those just extracted for non-punitive measures.

[14]When it comes to severity, preventative detention presupposes crimes against 'life, bodily integrity, personal freedom or sexual self-determination' (StGB §66).

[15]The rehabilitative dimensions of nonpunitive measures are not to be neglected, though, since they are sometimes appealed to help distinguish particularly severe instances of non-punitive measures from punitive sanctions. This is a particularly pressing issue when it comes to preventative detention, which may look and feel like punishment to the agent. Instructively, in Germany preventative detention was carried out similar to a jail sentence until in 2009 the European Court of Human Rights (*ECHR* 19359/04) objected to this practice by arguing that it made preventative detention resemble a jail sentence and thus punishment. The German Federal Constitutional Court (*BVerfGE* 128, 326) advised German legislature in 2011 to implement a suitable difference between punitive jail time and preventative detention, with one key distinguishing factor being that preventative detention is a 'freedom-oriented' rehabilitative measure (Frister 2023, n. 26–27; Meier 2019, §5.3.3.1). A similar concern with the rehabilitative dimension of non-punitive measures has also been noted by the US Supreme Court as a distinguishing factor to punitive sanctions (cf. *United States v. Halper*).

https://doi.org/10.1017/epi.2025.10063 Published online by Cambridge University Press

4. Non-punitive epistemic measures

Following Boult (2024c), I will assume a broadly communal metaepistemology as a backdrop for my discussion. Recall that on such a view, the purpose of our assessments and responses to epistemic conduct is, broadly conceived, to facilitate the epistemic division of labor and protect the epistemic commons. We collectively care about epistemic goods being widely available in our epistemic community, and we distribute them via relations of mutual epistemic trust, expecting each other to both be and trust reliable informants.

What role do epistemic trust reductions play in such a view? My proposal is that they should be understood as NoPEMs. They exclusively serve the purpose of *protecting* the epistemic community and the epistemic commons from unreliable informants. Reducing trust in those who often and expectedly fail epistemically is a form of *prevention*: it prevents unreliable informants from spreading badly formed beliefs within the community. It is easy to see how: if epistemic trust is what allows information to flow between agents, then depriving unreliable informants of this currency is a way of preventing them from epistemically affecting others and our epistemic goods negatively.

My proposal receives substantial support from the fact that the extracted appropriateness conditions for non-punitive measures also straightforwardly hold for trust reductions. I will demonstrate this in what follows.

(1) Trust reductions are responses to *epistemic failings*, that is, negatively evaluated epistemic conduct. As mentioned earlier, this includes not just badly formed beliefs, but also instances of bad reasoning or improper epistemic functioning, and so on.

(2) As Fricker (2007 §2.3, §6) has shown in detail, trust reductions undoubtedly impose negative epistemic consequences on the agents, which are ordinarily considered to be unpleasant. Trust reductions are applicable irrespective of culpability. In Fortune, Stephen is culpable for his believing based on the fortune-telling, and it is appropriate to reduce trust in him. In Senile Mother, the mother non-culpably believes the scammer. While her senility is a good excuse (or, depending on the severity, even an exemption), it is still appropriate to reduce trust in her.

(3) RMA's characterizations of the requirements for trust reductions to be appropriate read straightforwardly like the expectation requirement of non-punitive measures. Here is a representative example:

> When someone genuinely impairs their epistemic relationship with another, they *reveal* themselves prone to culpably epistemically fail in a *wide-ranging set of circumstances*. Typical ways of impairing epistemic relationships involve culpably flawed *epistemic dispositions* (e.g., dogmatism, proneness to wishful thinking, intellectual laziness). When someone has simply made an honest mistake or has a good excuse for their epistemic failing, this need not entail that they are prone to culpably epistemically fail in a wide-ranging set of circumstances. (Boult 2024b, 394, my emphases)

It is hard to not read this passage as asking for a 'symptomatic relation' between one's initial epistemic failings and one's epistemic character and dispositions, where the initial epistemic failing *reveals* something epistemically bad about one's epistemic dispositions and character.[16] What bad is revealed, in turn, needs to ground a (reasonable) expectation of *sufficiently likely* future epistemic failings, that is, failings in a wide range of similar circumstances. I agree with Boult that this kind of expectation requirement has

[16]In the same vein, Schmidt (2024, §4.2—3) suggests that trust reductions are only appropriate responses to manifestations of epistemic vices.

https://doi.org/10.1017/epi.2025.10063 Published online by Cambridge University Press

to be met for trust reductions to be appropriate. It seems inappropriate to reduce trust in ordinary agents in response to an isolated epistemic failing. That appropriate trust reductions presuppose an expectation of future epistemic failings is explained by the preventative character distinctive of NoPEMs: only given such an expectation is a trust reduction plausibly a proportional way of balancing our collective epistemic interests with the individual's interests of being regarded as trustworthy.[17]

Furthermore, proponents of RMA rightly point out that trust reductions are only appropriate as long as the condition that grounds our expectation of future epistemic failings persists. For instance, Boult (2024c, p. 91) considers Barney who unknowingly visits fake barn country and thus a highly deceptive environment. As long as he is driving through barn country, we do well to reduce trust in him on barn-related matters. After Barney leaves fake barn country and enters real barn country again, we are inclined to reinstate our trust in him on these matters.

What is perhaps less obvious is whether there is also a requirement of *sufficient severity* of future epistemic failings. There are two ways of thinking about this. On the one hand, one might think that this constraint is restricted to the legal domain and has no proper analog constraint in the epistemic domain. Topic-restricted trust reductions are (arguably) less substantial and wide-reaching than interference with one's rights and liberties by the state.[18] On the other hand, one might think that there is an epistemic analog to sufficient severity. Intuitively, we are less stringent in our responses to bagatelle epistemic failings. For instance, we might not bother responding to a false belief that you inattentively formed about an office supplies advert. We are much more stringent if, for instance, agents frequently fail epistemically, if they believe in false conspiracy theories about public health, if they do not lend credence to experts or if they deny the validity of basic inference patterns. It thus strikes me as plausible that the severity of an epistemic failing is a function of the *amount* of expected false beliefs, of the *community's interest* in them, the failing's *practice-undermining capacity,* and their *epistemic centrality* (Quine 1953). One interesting consequence of this view is that whether we respond with a trust reduction to an epistemic failing depends not only on the way the belief was formed, but also on its *content*. To tie this back to earlier observations, perhaps, the appropriateness of NoPEMs is restricted to particular topics. While this idea merits further exploration, doing so is a task for another occasion.

Before identifying rehabilitative aspects of NoPEM, let's consider a possible worry for the proposed expectation condition, namely that trust reductions are sometimes also 'backward-looking' (Flores and Woodard 2023, p. 2557). For instance, if someone epistemically fails, we might not only reduce trust in beliefs they *will* form, but also in beliefs they have already formed. How does this square with the idea that trust reductions target *future* epistemic failings?

I think that trust reductions in already formed beliefs also have a preventative purpose. To see this, note that we not only *form* new beliefs, but also *maintain* or revise beliefs. Maintaining beliefs is an ongoing process that regularly takes place, depending on whether we acquire relevant evidence or revisit our beliefs for various other reasons.

[17]What about cases in which trust in someone is already minimal, and we cannot reduce it further, even though the expectation condition is met? Perhaps, the continued epistemic failings make the trust reductions more 'sticky' and us less inclined to consider an increase in trust any time soon. Perhaps not responding further is just right: we think that all preventative measures are already in place, so there is no need for a further trust reduction. Still, there must remain a baseline level of trust. Even if we have minimal trust in an agent in a psychiatric facility with a condition that makes them lose contact with reality, we still should lend credence to and investigate their claim that they are abused by their doctors or caretakers.

[18]But see Watson (2021) for the view that we have *epistemic* rights.

https://doi.org/10.1017/epi.2025.10063 Published online by Cambridge University Press

It is thus a continuing source for future epistemic failings. If Stephen has been believing for some time that he should buy Boeing stocks because they will 'take off' very soon, and if he has maintained this belief by dogmatically responding to counterevidence ever since, then a 'backward-looking' trust reduction rather aims at preventing the fallout of Stephen's future maintenance failures.[19]

(4) Let's check the final condition and elaborate on the subordinate rehabilitative dimension that is sometimes involved in non-punitive measures. Is there an epistemic analog here? Again, I think so. Trust reductions can be accompanied by responses that are conducive to rehabilitation. Consider Fortune. Amari could appropriately respond to Stephen's failing with understanding and curiosity ('Why do you trust a fortune-teller only on these matters, but not on others?'), respond with criticism to cultivate reason-responsiveness ('Trusting a fortune-teller is irrational!', cf. Piovarchy 2021), provide educational resources ('Here are some basics on cyclical market trends.') or encourage him to epistemically atone (Woodard 2024). More paternalistically, we might also offer assistance, such as tutoring or guardianship, or even restrict access to bad sources. If these rehabilitative efforts turn out to be ineffective, they are plausibly discontinued (see the discussion of §63 on p. 7), resulting in purely preventative instances of NoPEMs.[20,21]

Summing up, there is ample reason to think that trust reductions are NoPEMs: They are (1) responses to epistemic failings, (2) impose negative epistemic consequences, (3) don't generally presuppose culpability, and (4) are only appropriate if there is an expectation of sufficiently likely and (arguably) severe future epistemic failings and only as long as the condition grounding this expectation persists. (5) They also sometimes involve a rehabilitative dimension.[22]

[19]What about 'backward-looking' trust reductions in *dead* people (who are clearly not maintaining their beliefs anymore)? While I cannot extensively discuss this issue here, I share Boult's (2024c, pp. 105–7) skepticism that our responses to the dead are significantly similar enough to our responses to the living. While Boult (ibid.) has some trouble explaining in what sense epistemic relationships to the dead are sufficiently distinct from our other epistemic relationships, there is principled reason on my account for thinking that we don't reduce trust in the dead: measures are only for the living. With that being said, I think trust reductions do play a role in preventing the (continued) negative impact of the dead's epistemic failings: but they target the living that keep these epistemic failings *alive* and in circulation.

[20]For recent empirical work of the effectiveness on various approaches to epistemic rehabilitation, see Ecker *et al.* (2022); O'Mahony *et al.* (2023); Douglas *et al.* (2024).

[21]Believers who culpably epistemically fail might not always see themselves as culpably epistemically failing and, when it comes to issues important to them, might remain insensitive to provisions of countervailing evidence, epistemic criticism, and trust reductions. These responses may sometimes even contribute to a further entrenchment of the believers' steadfastness, particularly so if the latter is rooted in deep disagreement or due to uncooperative epistemic environments, like epistemic bubbles and echo chambers (see, e.g., Nguyen 2020, p. 146). Here, rehabilitative efforts might be ineffective and eventually be discontinued, so NoPEMs, in these cases, are plausibly reduced to their primary purpose, namely the prevention of the further spread of epistemically bad beliefs. Deep disagreement and uncooperative epistemic environments ultimately raise vexed issues, and the jury is still out on how best to address them. While there is an interesting relation to be explored between these phenomena and individualized preventative measures like NoPEMs, I have to leave doing so to future research. Thanks to an anonymous referee for inviting me to elaborate on this.

[22]It is an interesting question what other NoPEMs might be. Possible candidates might be taking away someone's opportunity to learn (just think of study program disenrollment after having failed too many times at passing a mandatory examination), taking away educational authority (think of a teacher losing their license because they believe and teach pernicious conspiracy narratives) or monitor and review epistemic conduct (think of a coder who, after producing too much erroneous code, has his code reviewed much more frequently). While these possible NoPEM candidates merit further discussion and exploration, I will in this paper focus on trust reductions as the paradigmatic NoPEM.

https://doi.org/10.1017/epi.2025.10063 Published online by Cambridge University Press

My account of trust reductions as NoPEMs is explanatorily powerful. It explains why responses to culpable epistemic failings often involve trust reductions: non-punitive measures accompany culpability-tracking responses[23] if the culpable epistemic failing we respond to grounds an expectation of future epistemic failings. The account also explains why we sometimes reduce trust in non-culpable agents: NoPEMs don't presuppose culpability and can be appropriate responses to both culpable and non-culpable epistemic failings. My account explains why our responses to epistemic failings can include positively valenced responses, like curiosity and understanding: positively valenced responses will sometimes be included in NoPEMs and are conducive to epistemically rehabilitating agents. Finally, the account explains why trust reductions are only appropriate if there is an expectation of future epistemic failings: the purpose of trust reductions is to protect the epistemic commons from future epistemic failings. Only given such an expectation do our collective epistemic interests proportionally outweigh the individual's epistemic interests of being regarded as someone we can epistemically trust.

5. NoPEMs, RMA, and the challenges

In what follows, I will put my account to work to assess RMA's challenges and extant responses to them. I will then take stock of the options remaining for RMA and consider alternatives.

Recall that RMA faces two challenges: it counterintuitively counts as epistemic blame positively valenced responses to culpable epistemic failings and trust reductions in non-culpable epistemic agents.

Regarding the first challenge, Boult (2020, p. 528; 2024c, p. 104) denies that positively valenced responses fit one's judgment that the agent in question is blameworthy. He thinks that in making this assessment, we can rely on a 'shared grasp of the responses we *typically associate* with blame and relationship impairment' (my italics). However, the discussion from the last section showed that we cannot. Different kinds of responses, namely culpability-tracking responses (such as epistemic blame) and non-punitive measures, are often run together, namely in cases in which someone's epistemic failing is both culpable and grounds an expectation of future epistemic failings. Given a sufficiently rich conceptualization of our responses to epistemic failings, typical association is insufficient for a response to be a paradigmatic, constitutive component of epistemic blame.

Another way to address the first challenge is to modify RMA's account of epistemic blame to simply exclude positively valenced responses (Boult 2024c, pp. 72–73):

> **Epistemic Blame′** One's response to an agent *S*'s epistemic failing constitutes epistemic blame iff it is constituted by a judgment of blameworthiness, plus a negatively valenced modification to one's epistemic relationship that is a fitting type of response to that judgment.

However, as Boult is acutely aware (ibid.), this modification seems very much ad hoc.

My account provides a principled motivation on RMA's behalf for the exclusion of positively valenced responses in Epistemic Blame′. Recall that non-punitive measures

[23]Of course, I haven't specified yet what I take to be culpability-tracking epistemic responses (which we may or may not call 'epistemic blame'). I will consider various plausible candidates later on. To avoid getting side-tracked right now, I will continue with a minimal schematic conception of culpability-tracking epistemic responses: they are only appropriate if they are responses to culpable epistemic failings.

https://doi.org/10.1017/epi.2025.10063 Published online by Cambridge University Press

sometimes involve rehabilitative efforts. I argued above that positively valenced responses, such as showing understanding and curiosity, are naturally understood as being rehabilitative. To substantiate this further, recent empirical data suggest that being empathetic and receptive to those who, for instance, hold false conspiracy beliefs is rehabilitative: it can be an effective way of gaining their confidence and ultimately reducing their proneness to conspirational thinking.[24] My account classifies positively valenced responses to epistemic failings as being part of the rehabilitative dimension of NoPEMs.

Interestingly, Boult seems sympathetic to two-pronged responses of this kind: about a case in which a good friend drifted off into a YouTube conspiracy rabbit hole, Boult (2024c, p. 185f, also p. 191) suggests that we both reduce trust in our friend *and* respond in a positively valenced fashion by 'warming' our friendship with him to prevent him from further drifting off. However, Boult thinks that the 'warming' is a distinct, *non-epistemic* response concerned with one's *friendship* (ibid., p. 186). My worry with Boult's response is that it does not generalize, since it is easy to imagine cases in which understanding or curiosity is fitting, but not as a 'warming' of the relationship. It is an advantage of my proposal that it offers an epistemically unified account of our responses in these kinds of cases and, more generally, a principled motivation for the exclusion of positively valenced responses in Epistemic Blame′: positively valenced responses belong to NoPEMs, not epistemic blame.[25]

In response to the second challenge, Boult (2024b, c) has argued that trust reductions in excused agents lack a crucial condition for being manifestations of epistemic blame, namely 'meaning'. Drawing on Scanlon's work, Boult suggests that 'meaning' is tied to the 'agent's intention' (Boult 2024c, p. 65) and is characterized as the 'significance [epistemic conduct] has for oneself and the other people with whom one stands in various sorts of relationships' (ibid.). Note that 'meaning' can't just be bad epistemic consequences, since the epistemic failings of excused agents also have bad epistemic consequences for themselves and others. Rather, the 'meaning' of a culpable epistemic failing 'reveals [something epistemically bad] about how the agent regards their epistemic relations with others' (Boult 2024b, p. 394).

A general concern with this proposal is that it seems to make an epistemic failing a mere *evidential means* for what Boult claims to be the primary difference-maker of blameworthiness, viz., the agent's regard for their epistemic relations to others. A consequence of this view is that epistemic failings are, in principle, dispensable when it comes to assigning epistemic blame. Suppose, for instance, that there were other ways to find out about whether an agent disregards their epistemic relationships (say, via a brain scanner, via a search of their phone, and so on). If we could learn in these ways that they do, it counterintuitively would seem fitting to respond to their disregard of epistemic relationships with epistemic blame, even if no epistemic failing has occurred. Thus, the epistemic failings are mere evidential means for revealing how agents regard their epistemic relationships which sits uneasily with the natural idea that epistemic blame is primarily a response to *epistemic failings*.[26]

[24]See, e.g., Ecker *et al.* (2022, p. 22) and Douglas *et al.* (2024, §7). See both and O'Mahony *et al.* (2023) for the extensive literature on epistemic rehabilitation, which shows that both preemptive (e.g., 'prebunking' inoculation) and ex post approaches can be effective.

[25]This is not to deny that *negatively valenced* responses can also have rehabilitative effects. However, they are not categorized as non-punitive, since it is inappropriate to respond in a negatively valenced fashion to non-culpable agents.

[26]That epistemic blame is *primarily* a response to epistemic failings is compatible with Boult's observation that, in some cases, we may be epistemically blameworthy for our ill-adjusted epistemic expectations toward others (Boult 2024c, §5.3.1).

https://doi.org/10.1017/epi.2025.10063 Published online by Cambridge University Press

A related general concern is that Boult's proposal seems to get the relation between epistemic failings and how one regards one's epistemic relationships backward. Just like stealing *makes* you a thief, believing recklessly *makes* you a reckless believer, who disregards his epistemic relationships by not meeting the expectation of taking due care in believing. In the same vein, if someone is disposed to steal but never manifests this disposition, it seems intuitively incorrect to call him a thief; at most, we might call him a *potential* thief (Alston 1970, p. 26; Duff 1996, p. 189). The claim that the link between epistemic failings and an agent's regard for their epistemic relationships is merely evidential does not seem to capture that epistemic failings are plausibly constitutively prior to how failing agents regard their epistemic relationships.[27]

A more particular worry concerns the 'meaning'-criterion. While I find it quite hard to pin down Boult's exact view, I think he offers at least two conceptions of 'meaning', neither of which yields a plausible criterion. On a first reading, culpable epistemic failings involve bad epistemic intentions. When discussing a case in which an agent *deliberately* takes a pill that makes them reason hastily and dogmatically, Boult (2024c, p. 80) argues that the agent has (culpably) impaired his epistemic relationships. If, however, he was tricked into taking the pill, he would have a good excuse and would not have impaired his epistemic relationship (ibid.). On this *intention reading* of 'meaning', having epistemically bad intentions when one epistemically fails is thought to be a crucial condition for being epistemically blameworthy. One might think, for instance, that by deliberately taking the pill, one accepts that one will be an unreliable informant afterward, thereby revealing a disregard towards one's epistemic relationships, impairing them.[28]

The intention reading does not yield a plausible 'meaning'-criterion. 'Meaning' of this kind is not necessary: when we culpably epistemically fail, we rarely will have bad intentions toward our epistemic relationships.[29] This is because when we form beliefs, we are not typically concerned with our epistemic relationship. At most, we intend to adjust our beliefs to our evidence. When we form non-reflective, 'animal beliefs' (cf. Sosa 2009), for example, perceptual beliefs, we may not intend anything at all. But even in cases of epistemic failings where we have intentions that concern our epistemic relationships, it's not that we intend to (or take ourselves to) impair them. Instead, we will often erroneously think that we are doing well epistemically.[30]

Bad intentions are also not sufficient for blameworthiness. To see this, consider the following case:

> **Contrarian** Caleb is a habitual contrarian who always wants to believe the opposite of what others believe. Suppose, however, that his perceptual apparatus functions perfectly well and, against the considerable but insufficient force of his bad intentions, has resulted in him ordinarily acquiring true beliefs in normal lighting

[27]The raised concerns resemble similar concerns about kindred variants of character-based views of excuse and culpability (see, e.g., Duff 1996, ch. 7; Moore 2010, ch. 13, for discussion).

[28]I think the intention reading should not be understood as the claim that the agent, for their epistemic conduct to be blameworthy, needs to intend to impair their epistemic relationships. Rather, an intention needs to be involved that is (in one way or another) epistemically inappropriate. For instance, if *A* intends to find out whether *p*, but lazily cuts corners in his inquiry and forms a belief that *p* based on insufficient evidence, then *A*'s intention doesn't aim at impairing epistemic relationships, but is still epistemically inappropriate in the relevant sense.

[29]See also Piovarchy (forthcoming, §3) for this worry and for further excellent critical discussion of Boult's account and 'meaning'- criterion.

[30]A classic illustration of this is the Dunning–Kruger effect: low competence in, for instance, reasoning tasks is linked to an overconfidence in one's ability to perform them (Kruger and Dunning 1999).

https://doi.org/10.1017/epi.2025.10063 Published online by Cambridge University Press

conditions. Unbeknownst to Caleb, however, he has just entered a highly deceptive environment and forms many false perceptual beliefs as a result.

Intuitively, Caleb is still excused (or even justified, depending on your view about epistemic justification) and thus not blameworthy for believing falsely, despite his bad intentions. Contrarian is thus a case in which an excusing (or justifying) condition *normatively screens off* bad intentions. Hence, bad intentions are not sufficient for blameworthiness.

A second reading of 'meaning' might be called the *attitude and conduct* reading: Drawing on Smith (2005), Boult (2024c, p. 111) notes that 'attitudes and actions that we do not (seem to) have control over can nevertheless express a great deal about how we regard one another'. Likewise, doxastic responses are often involuntary and can reveal a lot about our epistemic character and how we regard our epistemic relationships. For instance, believing the contents of a tabloid newspaper reveals something bad about one's epistemic character (Boult 2024c, p. 112), namely that one is gullible. But doing so after a gullibility pill is slipped into one's coffee does not.

The resulting 'meaning'-criterion is not plausible either. 'Meaning' in this sense is not necessary for blameworthiness. We sometimes epistemically fail due to *slips* (Williamson forthcoming; Heil 2022; cf. Amaya 2011) caused by stupidity, clumsiness, or no discernible reason at all. Suppose you forget to congratulate your friend on her birthday because you made the stupid mistake of failing to transfer the date while reorganizing your calendar. Or suppose you clumsily drop a pile of newspapers, pick out the wrong one, and form a false belief about today's winning lottery numbers. In both cases, you intuitively are blameworthy for your epistemic failing.[31] While some one-off epistemic failings may be due to lack of care or other defects of character, it is not plausible that they always (or even often) are. Hence, there are cases of culpable epistemic failings that don't reveal anything about your character or how you regard your epistemic relationships.[32,33]

[31]Interestingly, Boult (2023, p. 818) seems to agree that epistemic failings caused by stupid mistakes can be blameworthy.

[32]This parallels the worry in legal theory that character-based accounts of excuses struggle to account for isolated instances of negligent conduct (Moore 2010, pp. 590–1). See also the related discussion on slips (sometimes called 'unwitting omissions' or 'performance failures') in ethics (e.g., Amaya and Doris 2014; Rudy-Hiller 2019; Murray and Vargas 2020).

[33]On behalf of Boult, one might respond that in cases of slips, the agents might be *less* epistemically blameworthy, and propose that the *degree of blameworthiness* corresponds to *how much an agent's epistemic failing reveals* about their disregard towards their epistemic relationships. While I share the intuition that slipping agents are typically less blameworthy than agents whose epistemic conduct reveals bad intentions or gross negligence, the proposal is not able to vindicate this verdict. Generally speaking, slips (at least if properly distinguished from negligent errors and other mistakes; see Amaya 2011) need not reveal *anything* about one's character or about how one regards others (cf. Amaya and Doris 2014, p. 263). Agents who slip typically have sufficient control of their φ-ing, recognize, intend to act on, and are motivated by, their reasons for φ-ing, yet small lapses (e.g., in concentration or memory) might prevent them from registering factors relevant to correct φ-ing or subvert the successful execution of the intention to φ. Consider 'death by hyperthermia' cases, in which parents who are on their way to work unwittingly forget their small children in their cars, often with fatal outcomes. While these cases sometimes happen due to gross negligence, they also regularly happen due to slips by parents who are loving, attentive, and well-organized (Collins 2006). While the latter are arguably blameworthy to a lesser degree, the abovementioned proposal would predict that, if slips don't reveal anything about our character or attitudes toward others, the appropriate degree of blameworthiness is not just less, but *nil*. This is a highly revisionary assessment, given that ordinary and legal practice show that slipping agents are regularly considered blameworthy to a significant degree (Murray *et al.* 2019). Finally, perhaps tellingly, Boult himself opts for a different view on degrees of epistemic

https://doi.org/10.1017/epi.2025.10063 Published online by Cambridge University Press

'Meaning' in this second sense is also not sufficient. This can be seen as soon as we look beyond 'environmental or exogenous' (Boult 2024b, p. 394) excusing condition. Sometimes, excusing conditions are both grounded in one's epistemic character *and* affect one's epistemic relationships by violating legitimate expectations. Consider the following case:

> **Anxious Volunteer** It's Alfred's first day of volunteering at the neighborhood soup kitchen. His job is to welcome the arriving guests and pass on their individual needs to the workers in the back. It's a fairly slow day, and there is only a small crowd waiting outside. Yet, the crowd is large enough for Alfred to learn quickly and to his own surprise, that whenever he interacts with the people waiting, he feels overwhelming anxiety and panic. Due to this debilitating emotional distress, Alfred becomes withdrawn and distrustful; when guests tell him about their needs, he often doesn't believe them. As a result, he forms many false beliefs about what has to be prepared in the back (namely, way too little, given what he was told by the guests). After asking to be transferred to another position, he connects his experience to past events and becomes aware that he has always been extremely socially anxious when engaging with crowds.

Alfred's anxiety is overwhelming and thus excuses his epistemic failings. Yet, his epistemic failings reveal something *epistemically bad* – his unfounded distrust – about how he regards his epistemic relationships to others. Furthermore, Alfred violates the legitimate, default expectations of the guests, his co-workers, and himself to handle these fairly ordinary circumstances well and be trusting and trustworthy concerning the needs communicated to him. Still, even though his epistemic failing reveals something epistemically bad about him and violates epistemic expectations, Alfred is excused.

Consider two possible replies: First, Boult (2023, p. 818; 2024b, p. 393) might reply that 'good epistemic relationships involve agents expecting others to avoid epistemic failings, unless they have a legitimate excuse, or perhaps some overriding reason not to'. On this view, excused epistemic failings could not in principle violate epistemic expectations, because excusing conditions are simply baked into our expectations. However, there is a perfectly reasonable sense in which the default expectation towards Alfred is to do well in the circumstances at issue. That we excuse him for violating these expectations is a 'concession to human frailty', as legal theorists sometimes put it, not a reason to believe that we didn't expect him to do well in the first place.[34]

In the light of these difficulties, I think proponents of RMA would do well to embrace my account of trust reductions and give up on the idea that trust reductions are typical constitutive components of epistemic blame. My account offers a principled response to RMA's first challenge: positively valenced responses belong to NoPEMs, not to epistemic blame. My account also resolves the second challenge to RMAs: if trust reductions

blameworthiness. In a recent paper (Boult 2024a), he proposes an account of degrees of epistemic 'criticisability' – noting his sympathies for interpreting criticizability in terms of blameworthiness (ibid., pp. 433–4) – which is control-based and in which the relevant parameter is the difficulty in doing otherwise (ibid., p. 438). Whatever its merits, these features make it differ substantially from the conception of epistemic blame underlying RMA, and thus of no use for saving the appeal to 'meaning'. I am grateful to an anonymous referee for encouraging me to discuss degrees of epistemic blameworthiness in this context.

[34]See also Piovarchy (forthcoming, §3) for related worries. Still, the interaction between excuses and expectations is a substantial and difficult issue that is beyond the scope of this paper. The same is true of one promising diagnosis in this regard, namely Hazlett's (2024) recent proposal that ties epistemic expectations much more intimately to epistemic shame, not epistemic blame.

https://doi.org/10.1017/epi.2025.10063 Published online by Cambridge University Press

merely *accompany* epistemic blame responses, but don't *partially constitute* them, then no worry about reductions of trust in non-culpable agents arises.

6. Taking stock

Where does this leave RMA? Embracing my account in order to parry the two challenges might seem like a pyrrhic victory for proponents of RMA because without trust reductions, RMA's account of epistemic blame looks fairly anemic. Proponents of RMA could counter this impression by offering a richer account of epistemic relationship-modifications. They might identify more promising candidates for blame-constituting responses to epistemic failings or claim that the latter are simply varied and contextually determined (Boult 2024c, p. 88). In any case, I think proponents of RMA have their work cut out for them.

Those not wedded to RMA might take my argument to be grounds for skepticism or outright rejection of RMA. After all, the latter has partly been motivated by the idea that trust reductions are paradigmatic ways of altering epistemic relationships and thus partially constitutive of epistemic blame. However, if my account is correct, then trust reductions are, in fact, not partially constitutive of epistemic blame.

If we reject RMA, one option would be to start with my account of trust reductions as NoPEMs and enrich it with an account of culpability-tracking responses to epistemic failings. While this is ultimately a project for another paper, let me briefly sketch some options.

First, there are various candidates in the literature on epistemic blame that we might appeal to (see fn. 2). It might indeed be worth revisiting some of these competitors to RMA because one of the main complaints against them is that they have trouble explaining the putative 'coolness' of epistemic blame (Boult 2020, p. 519f). However, if my account is right, then proponents of RMA might be mistaken in thinking that epistemic blame is particularly 'cool', at least if they think so because they mistakenly take 'cool' trust reductions to be instances of epistemic blame.

Second, one might think that the notion of blame belongs to the moral and legal domain only (Kauppinen 2018, p. 2) and that we should thus look for accounts of culpability-tracking responses that do without the notion of epistemic blame. In this regard, I am sympathetic to Dogramaci's view (2012, 2015) that the purpose of epistemic assessment is to identify reliable informants for deference and to facilitate the communal coordination regarding the use of reliable belief-forming methods.[35] In response to culpable epistemic failings, say, a belief formed by wishful thinking, we respond with a negative assessment ('Your belief is irrational!'), thereby expressing disapprobation to influence our audience to not use the employed belief-forming method and to not defer to the thus-formed belief. Dogramaci's account of culpability-tracking responses would complement my account nicely: trust reductions as NoPEMs would be the right response to agents who can be expected to form false beliefs in the future because they are culpably or nonculpably *unreceptive* to the force of our epistemic assessments. While I think this sketch is highly promising, I will have to leave filling in the details for another occasion.

7. Conclusion

In this paper, I proposed a novel account of epistemic trust reductions and employed it to assess the RMA of epistemic blame and its challenges. I argued that proponents of

[35]See as well Greco and Hedden (2016). For discussion, see Dogramaci and Horowitz (2016); Daoust (2017), Thorstad (2019), Horowitz *et al.* (2024); Heil (Ms).

https://doi.org/10.1017/epi.2025.10063 Published online by Cambridge University Press

RMA have misidentified epistemic trust reductions as paradigmatic instances of epistemic blame and that epistemic trust reductions should rather be understood as non-punitive epistemic measures. My account of epistemic trust reductions as NoPEMs is not only explanatorily powerful but also provides a principled response to the two challenges that RMA faces. However, I have argued that embracing my account comes with its own challenge for proponents of RMA. They have to offer alternative candidates for epistemic blame responses. Those pessimistic about the prospects of meeting this challenge might choose another starting point altogether, one of which I have defended in this paper.[36]

References

Alston W.P. (1970). 'Toward a Logical Geography of Personality: Traits and Deeper Lying Personality Characteristics.' In H.E. Kiefer and M.K. Munitz (eds), *Mind, Science, and History*, pp. 59–92. Albany, NY: State University of New York Press.

Amaya S. (2011). 'Slips.' *Noûs* **47**(3), 559–76.

Amaya S. and Doris J.M. (2014). 'No Excuses: Performance Mistakes in Morality.' In J. Clausen and N. Levy (eds), *Springer Handbook of Neuroethics*, pp. 253–72. Dordrecht: Springer.

Austin J. (1957). 'A Plea for Excuses.' *Proceedings of the Aristotelian Society* **57**, 1–30.

Ballantyne N. (2019). 'Epistemic Trespassing.' *Mind* **128**(510), 367–95.

Boult C. (2020). 'There is a Distinctively Epistemic Kind of Blame.' *Philosophy and Phenomenological Research* **103**(3), 518–34.

Boult C. (2021a). 'Epistemic Blame.' *Philosophy Compass* **16**(8), e12762.

Boult C. (2021b). 'Standing to Epistemically Blame.' *Synthese* **199**(3–4), 11355–75.

Boult C. (2023). 'The Significance of Epistemic Blame.' *Erkenntnis* **88**(2), 807–28.

Boult C. (2024a). 'Degrees of Epistemic Criticizability.' *Philosophical Quarterly* **74**(2), 431–52.

Boult C. (2024b). 'Epistemic Blame as Relationship Modification: Reply to Smartt.' *Philosophical Studies* **181**, 387–96.

Boult C. (2024c). *Epistemic Blame: The Nature and Norms of Epistemic Relationships*. Oxford: Oxford University Press.

Boult C. (2024d). 'The Relational Foundations of Epistemic Normativity.' *Philosophical Issues* **34**(1), 285–304.

Brown J. (2018). 'What is Epistemic Blame?' *Noûs* **54**(2), 389–407.

Brown J. (2020). 'Epistemically Blameworthy Belief.' *Philosophical Studies* **177**(12), 3595–614.

Chislenko E. (2020). 'Scanlon's Theories of Blame.' *Journal of Value Inquiry* **54**(3), 371–86.

Chrisman M. (2022). *Belief, Agency, and Knowledge*. Oxford: Oxford University Press.

Collins J.M. (2006). 'Crime and Parenthood: The Uneasy Case for Prosecuting Negligent Parents.' *Northwestern University Law Review* **100**, 807.

Craig E. (1990). *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*. Oxford: Oxford University Press.

Daoust M. (2017). 'Epistemic Uniqueness and the Practical Relevance of Epistemic Practices.' *Philosophia* **45**(4), 1721–33.

Dogramaci S. (2012). 'Reverse Engineering Epistemic Evaluations.' *Philosophy and Phenomenological Research* **84**(3), 513–30.

Dogramaci S. (2015). 'Communist Conventions for Deductive Reasoning.' *Noûs* **49**(4), 776–99.

Dogramaci S. and Horowitz S. (2016). 'An Argument for Uniqueness about Evidential Support.' *Philosophical Issues* **26**(1), 130–47.

[36]I would like to thank Cameron Boult, Jochen Briesen, Antti Kauppinen, Tim Smartt, Adam Piovarchy, Susanna Schellenberg, Michael Vollmer, Alexandra Zinke, an anonymous referee of this journal, and Christopher von Bülow for their valuable comments. For stimulating and helpful feedback, I am grateful to audiences at the Epistemic Accountability workshop at the University of Notre Dame, Sydney, the Goethe Epistemology Meeting 2024 at Goethe University Frankfurt, the Blame, Excuses and Responsibility conference at the University of Neuchâtel and the Institutional Epistemology workshop at the University of Helsinki.

https://doi.org/10.1017/epi.2025.10063 Published online by Cambridge University Press

Douglas K.M., Sutton R.M., Biddlestone M., Green R. and Toribio-Flórez D. (2024). 'Engaging with Conspiracy Believers.' *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-024-00741-0.
Dressler J. (2015). *Understanding Criminal Law*, 7th edn. Durham, NC: Carolina Academic Press.
Dubber M. and Hörnle T. (2014). *Criminal Law: A Comparative Approach*. Oxford: Oxford University Press.
Duff R.A. (1996). *Criminal Attempts*. Oxford: Oxford University Press.
Dyke M.M. (2021). 'Could Our Epistemic Reasons be Collective Practical Reasons?' *Noûs* **55**(4), 842–62.
Ecker U.K., Lewandowsky S., Cook J., Schmid P., Fazio L.K., Brashier N., Kendeou P., Vraga E.K. and Amazeen M.A. (2022). 'The Psychological Drivers of Misinformation Belief and Its Resistance to Correction.' *Nature Reviews Psychology* **1**(1), 13–29.
Fleisher W. (2025). 'Epistemic Practices: A Unified Account of Epistemic and Zetetic Normativity.' *Noûs* **59**(1), 289–314.
Flores C. and Woodard E. (2023). 'Epistemic Norms on Evidence-Gathering.' *Philosophical Studies* **180**(9), 2547–71.
Fricker M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.
Frister H. (2023). *Strafrecht Allgemeiner Teil*, 10th edn. Munich: C.H. Beck.
Goldberg S. (2017). 'Should Have Known.' *Synthese* **194**(8), 2863–94.
Goldberg S. (2018). *To the Best of Our Knowledge: Social Expectations and Epistemic Normativity*. Oxford: Oxford University Press.
Goldberg S. (2020). 'Trust and Reliance.' In J. Simon (ed), *The Routledge Handbook of Trust and Philosophy*, pp. 97–108. London: Routledge.
Graham P. (2015). 'Epistemic Norms as Social Norms.' In D. Henderson and J. Greco (eds), *Epistemic Evaluation: Purposeful Epistemology*, pp. 247–73. Oxford: Oxford University Press.
Greco D. (2021). 'Justifications and Excuses in Epistemology.' *Noûs* **55**(3), 517–37.
Greco D. and Hedden B. (2016). 'Uniqueness and Metaepistemology.' *Journal of Philosophy* **113**(8), 365–95.
Greco J. (2007). 'The Nature of Ability and the Purpose of Knowledge.' *Philosophical Issues* **17**(1), 57–69.
Hannon M. (2013). 'The Practical Origins of Epistemic Contextualism.' *Erkenntnis* **78**(4), 899–919.
Hannon M. (2015). 'Stabilizing Knowledge.' *Pacific Philosophical Quarterly* **96**(1), 116–39.
Hannon M. (2019). *What's the Point of Knowledge? A Function-First Epistemology*. Oxford: Oxford University Press.
Hannon M. and Woodard E. (forthcoming). 'The construction of epistemic normativity.' *Philosophical Issues*.
Hawley K. (2014). 'Trust, Distrust and Commitment.' *Noûs* **48**(1), 1–20.
Hazlett A. (2024). 'From Doxastic Blame to Doxastic Shame.' *Ergo: An Open Access Journal of Philosophy* **11**, 38.
Heil R. (2022). 'Finding Excuses for J = K.' *Thought. A Journal of Philosophy* **11**(1), 32–40.
Henderson D. (2009). 'Motivated Contextualism.' *Philosophical Studies* **142**(1), 119–31.
Henderson D. (2011). 'Gate-Keeping Contextualism.' *Episteme* **8**(1), 83–98.
Hieronymi P. (2004). 'The Force and Fairness of Blame.' *Philosophical Perspectives* **18**(1), 115–48.
Horowitz S., Dogramaci S. and Schoenfield M. (2024). 'Are You Now or Have You Ever been an Impermissivist? – A Conversation among Friends and Enemies of Epistemic Freedom.' In B. Roeber, M. Steup, E. Sosa and J. Turri (eds), *Contemporary Debates in Epistemology*. Hoboken, NJ: Wiley-Blackwell.
Kauppinen A. (2018). 'Epistemic Norms and Epistemic Accountability.' *Philosophers' Imprint* **18**, 1–16.
Kauppinen A. (2023). 'The Epistemic vs. the Practical.' In R. Shafer-Landau (ed), *Oxford Studies in Metaethics*, vol. **18**, pp. 137–62. Oxford: Oxford University Press.
Kruger J. and Dunning D. (1999). 'Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments.' *Journal of Personality and Social Psychology* **77**(6), 1121–34.
Kusch M. and McKenna R. (2020). 'The Genealogical Method in Epistemology.' *Synthese* **197**(3), 1057–76.
McGrath M. (2015). 'Two Purposes of Knowledge-Attribution and the Contextualism Debate.' In D.K. Henderson and J. Greco (eds), *Epistemic Evaluation: Purposeful Epistemology*, pp. 138–58. Oxford: Oxford University Press.
McHugh C. (2012). 'Epistemic Deontology and Voluntariness.' *Erkenntnis* **77**(1), 65–94.
Meier B.-D. (2019). *Strafrechtliche Sanktionen*. Berlin: Springer.
Moore M.S. (2010). *Placing Blame: A Theory of the Criminal Law*. Oxford: Oxford University Press.

https://doi.org/10.1017/epi.2025.10063 Published online by Cambridge University Press

Murray S., Murray E.D., Stewart G., Sinnott-Armstrong W. and Brigard F.D. (2019). 'Responsibility for Forgetting.' *Philosophical Studies* **176**(5), 1177–201.
Murray S. and Vargas M. (2020). 'Vigilance and Control.' *Philosophical Studies* **177**(3), 825–43.

Neta R. (2006). 'Epistemology Factualized: New Contractarian Foundations for Epistemology.' *Synthese* **150**(2), 247–80.
Nguyen C.T. (2020). 'Echo Chambers and Epistemic Bubbles.' *Episteme* **17**(2), 141–61.
Nottelmann N. (2007). *Blameworthy Belief: A Study in Epistemic Deontologism*. Dordrecht: Springer.
O'Mahony C., Brassil M., Murphy G. and Linehan C. (2023). 'The Efficacy of Interventions in Reducing Belief in Conspiracy Theories: A Systematic Review.' *PLoS One* **18**(4), e0280902.
Piovarchy A. (2021). 'What Do We Want from a Theory of Epistemic Blame?' *Australasian Journal of Philosophy* **99**(4), 791–805.
Piovarchy A. (forthcoming). 'Epistemic blame isn't relationship modification.' *Philosophical Quarterly*.
Quine W.V.O. (1953). *From a Logical Point of View*. Cambridge, MA: Harvard University Press.
Rettler L. (2018). 'In Defense of Doxastic Blame.' *Synthese* **195**(5), 2205–26.
Reynolds S.L. (2002). 'Testimony, Knowledge, and Epistemic Goals.' *Philosophical Studies* **110**(2), 139–61.
Rudy-Hiller F. (2019). 'Give People a Break: Slips and Moral Responsibility.' *Philosophical Quarterly* **69**(277), 721–40.
Rysiew P. (2012). 'Epistemic Scorekeeping.' In J. Brown and M. Gerken (eds), *Knowledge Ascriptions*, pp. 270–93. Oxford: Oxford University Press.
Scanlon T. (2008). *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, MA: Belknap Press of Harvard University Press.
Schmidt S. (2024). 'Epistemic Blame and the Normativity of Evidence.' *Erkenntnis* **89**(1), 1–24.
Sher G. (2005). *In Praise of Blame*. Oxford: Oxford University Press.
Smartt T. (2023). 'Scepticism about Epistemic Blame.' *Philosophical Studies* **180**, 1813–28.
Smith A. (2013). 'Moral Blame and Moral Protest.' In D.J. Coates and N.A. Tognazzini (eds), *Blame: Its Nature and Norms*. Oxford: Oxford University Press.
Smith A.M. (2005). 'Responsibility for Attitudes: Activity and Passivity in Mental Life.' *Ethics* **115**(2), 236–71.
Sosa E. (2009). *Reflective Knowledge*. Oxford: Oxford University Press.
Thompson C. (2017). 'Trust Without Reliance.' *Ethical Theory and Moral Practice* **20**(3), 643–55.
Thorstad D. (2019). 'Permissive Metaepistemology.' *Mind* **128**(511), 907–26.
van Gemmeren G. (2020). 'StGB § 63 Unterbringung in Einem Psychiatrischen Krankenhaus.' In V. Erb and J. Schäfer (eds), *Münchener Kommentar zum StGB*. Munich: C.H. Beck.
Vargas M. (2013). *Building Better Beings: A Theory of Moral Responsibility*. Oxford: Oxford University Press.
Watson L. (2021). *The Right to Know: Epistemic Rights and Why We Need Them*. London: Routledge.
Williams B. (1973). *Problems of the Self*. Cambridge: Cambridge University Press.
Williamson T. (forthcoming). 'Justifications, Excuses, and Sceptical Scenarios.' In J. Dutant and F. Dorsch (eds), *The New Evil Demon*. Oxford: Oxford University Press.
Wood D. (2010a). 'Punishment: Consequentialism.' *Philosophy Compass* **5**(6), 455–69.
Wood D. (2010b). 'Punishment: Nonconsequentialism.' *Philosophy Compass* **5**(6), 470–82.
Woodard E. (2024). 'Epistemic Atonement.' In R. Shafer-Landau (ed), *Oxford Studies in Metaethics*, vol. **18**, pp. 163–90. Oxford: Oxford University Press.

Roman Heil is a postdoctoral researcher at Goethe University Frankfurt. He gained his PhD from the University of Hamburg. He is the author of *Knowledge and Rational Action. Why What We Know Matters for the Rationality of What We Do* (2025, Routledge). His research focuses on topics in epistemology and the theory of practical rationality, with a recent emphasis on the intersection between epistemic and legal normativity.

Cite this article: Heil R. (2025). "Trust Reductions, Epistemic Blame, and Preventative Measures." *Episteme* 1–20. https://doi.org/10.1017/epi.2025.10063

https://doi.org/10.1017/epi.2025.10063 Published online by Cambridge University Press