




ARTICLE

Maintaining Reprehensibility for Epistemic Vice: Responsibility for Implicit Bias as Non-vicious Conduct

Carline Julie Francis Klijman^{1,2} 

¹Utrecht University, Utrecht, Netherlands and ²Eindhoven University of Technology, Eindhoven, Netherlands
Email: c.klijman.ac@gmail.com

(Received 16 March 2024; revised 21 June 2024; accepted 13 August 2024)

Abstract

Heather Battaly has argued that vice-epistemology has a Responsibility Problem. From analysing the ‘card-carrying feminist’ committing testimonial injustice due to implicit gender bias, Battaly argues that non-voluntarist vice-epistemologists are committed to either (1) counting some vices as blameworthy yet not reprehensible, or (2) holding agents equally responsible for cognitive defects as for implicit bias. This in turn implies that (2a) epistemic vices include certain cognitive defects or (2b) that implicit bias is excluded as epistemic vice. This paper aims to deflate the Responsibility Problem, by arguing that vice-epistemologists can embrace route (2b) without problematic implications. In applying Miranda Fricker’s ‘no-fault responsibility’ to the card-carrying feminist case, I defend the following three claims: First, Battaly’s analysis of the card-carrying feminist case is flawed, because it fails to acknowledge that the vice of testimonial injustice and implicit prejudice are conceptually distinct. Second, excluding implicit prejudice as a vice is actually compatible with Fricker’s theoretical commitments. Third, contrary to what much of the literature seems to assume, it isn’t that problematic to exclude (individual) implicit bias as a vice, or to assign equal responsibility for implicit prejudice as for cognitive defects like impaired vision.

Keywords: Vice epistemology; responsibility; implicit bias; testimonial injustice; epistemic blame

1. Introduction

Vice epistemology is concerned with questions about the nature, relationship to knowledge, and remedying of so-called epistemic vices, or epistemic features that reflect badly on the epistemic agent. The responsibilist tradition of vice-epistemology conceives of epistemic vices as character traits, dispositions or ways of thinking that involve some kind of motivation and hence, responsibility of the agent. Examples of such vices include closed-mindedness, epistemic negligence, epistemic arrogance, and epistemic servility. Responsibilist vice epistemology can struggle to deal with vices over which the agent seemingly has no control. An often-mentioned example is the closed-mindedness or dogmatic tendencies of the Hitler *Jugend*. These children have been indoctrinated such that they lacked control over the development of these vices. At the same time, we want to be able to criticize their problematic ways of thinking and ascribe some kind of

© 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.

responsibility. To account for these kinds of features as vices, whilst acknowledging the partial (lack of) control, vice-epistemologists have turned to non-voluntarist notions of responsibility as a way of holding people responsible for certain epistemic features even if they lack (full) voluntary control over them.

In an illuminating paper, Heather Battaly (2019) has argued that whilst it can account for the vices of the Hitler *Jugend*, the non-voluntarist notion of responsibility leads to other difficulties. Battaly primarily discusses Quassim Cassam's (2019) obstructionism to illustrate this challenge, but she maintains that it affects most if not all non-voluntarist views. Her analysis, which I will discuss in more detail below, applies obstructionism to the concept of implicit bias, and ultimately raises the following dilemma:

- (1) Either we have to give up the intuition that blameworthiness is a subset of reprehensibility, or:
- (2) We have to commit to holding epistemic agents equally responsible for their cognitive defects like impaired vision as for implicit bias. This in turn implies that (2a) we have to include cognitive defects as vices, or (2b) exclude implicit bias as a vice.

Neither option is satisfactory, leaving vice-epistemology with a Responsibility Problem. Battaly explicitly mentions Miranda Fricker's no-fault responsibility as a view that falls on the second horn of the dilemma, casting the net of responsibility too widely. This paper aims to deflate the worry over the Responsibility Problem, by showing that Fricker's no-fault responsibility is actually well equipped to deal with the dilemma given its theoretical commitments. In what follows, I will first reconstruct Battaly's Responsibility Problem in more detail, going over her analysis of implicit bias in the case of 'the card-carrying feminist'. I will then employ Fricker's no-fault responsibility to argue that vice epistemologists can embrace route (2b), excluding implicit bias as a vice, without giving up on any theoretical commitments, and without leading to implausible conclusions. My response to the Responsibility Problem follows from three main points:

First, I argue that Battaly's analysis of the card-carrying feminist is flawed, because she mistakenly equates implicit bias with testimonial injustice. Second, I argue that Fricker can bite the bullet and exclude implicit prejudice as a vice, without giving up on any theoretical commitments – including the classification of testimonial injustice as a vice. Third, I argue that the exclusion of implicit bias as a vice, as well as ascribing equal responsibility for implicit bias as for cognitive defects, are not as problematic or implausible as the literature assumes. In fact, it allows us to maintain the intuition that that blameworthiness is a subset of reprehensibility, offering obstructionism another solution to the Responsibility Problem. I also include some discussion on how my strategy compares to an earlier response to the Responsibility Problem from Alessandra Tanesini (2020).

2. The responsibility problem

Before elaborating on the challenge Battaly raises against vice-epistemologists who employ a non-voluntarist notion of responsibility, it is helpful to go over some key concepts that are crucial to her argument. Battaly employs a conception of epistemic vices from Cassam's obstructionism, according to which epistemic vices obstruct the acquisition or sharing of knowledge, and are 'blameworthy or otherwise reprehensible' (Cassam 2019: 23).

From the previous quotation, Battaly defers that blameworthiness is a subcategory of reprehensibility, i.e., not all reprehensible features are blameworthy, but blameworthy features are necessarily reprehensible. Let's unpack these terms further. According to Cassam, two conditions apply for some trait or way of thinking to be reprehensible: first,

it needs to be close enough to the agent that it reflects badly on them. Second, it needs to reflect a deep quality rather than a superficial one, referring to the kind of epistemic agent they are. In other words, reprehensible features are necessarily involved in epistemic *agency*. Two conditions need to be in place in order to hold someone blameworthy for their epistemic vices: 1) a harm condition (i.e. the vices lead to epistemic harms) and 2) a responsibility condition. Responsibility for epistemic vices can take two forms: *Acquisition responsibility*, meaning that one is responsible for acquiring the epistemic vice, or *revision responsibility*, meaning that one can be responsible for not addressing and correcting their epistemically vicious behaviour (Cassam 2019: 18).

These criteria for epistemic vices (obstructing knowledge acquisition and being blameworthy or otherwise reprehensible) seem plausible. However, so Battaly claims, upon further scrutiny they don't hold up in certain cases. More specifically, she argues for the following dilemma: obstructivism is committed to either counting some vices as blameworthy yet not reprehensible, or take agents as equally responsible for certain cognitive defects as for some vices (Battaly 2019: 29). This *Responsibility Problem*, as Battaly calls it, applies to all accounts of non-voluntarist responsibility for epistemic vices.

Battaly uses the case of the 'card carrying feminist' to illustrate this dilemma. Note that implicit bias and implicit prejudice are used interchangeably here. The case employs Miranda Fricker's now well-known concept of *testimonial injustice*; an instance whereby a speaker receives a credibility deficit due to identity-prejudice in the hearer (Fricker 2007: 4). The crucial point here is that testimonial injustice can be driven by *implicit prejudice*, even in the face of contradicting beliefs – actually, this is the more prominent form testimonial injustice takes. The card-carrying feminist holds explicit feminist beliefs, but nevertheless is influenced by implicit gender prejudice in their credibility judgement.

The dilemma Battaly (2019) sketches comes to light when we evaluate whether the implicit prejudice in this case constitutes a vice. It is clear that implicit prejudice can obstruct knowledge acquisition, i.e.; I take the harm condition to be in place. That means that the presence of either acquisition responsibility or revision responsibility would render the agent either acquisition blameworthy or revision blameworthy. If implicit prejudice is a vice, according to obstructivism, the card-carrying feminist must additionally be either acquisition blameworthy, revision blameworthy, or otherwise reprehensible for it. Implicit prejudices are practically impossible to prevent, as they result from our upbringing, socio-political environment and other contextual influences over which we have little control. Accordingly, acquisition blameworthiness quickly gets ruled out. However, given the extent to which implicit prejudice is a known phenomenon, as are unequal power relations, it seems fair to say that the card-carrying feminist has ample opportunity to correct their implicit prejudice, and hence it seems plausible that they are revision blameworthy.

So far, we have established that the card-carrying feminist lacks acquisition blameworthiness but has revision blameworthiness. Rests the question whether the card-carrying feminist is *reprehensible* for their implicit prejudice. And this is where Battaly's dilemma becomes clear. There are two ways to answer the question: on the one hand, one could argue that since the implicit gender bias contradicts the actual beliefs of the feminist, it *bypasses* their epistemic agency. It doesn't reflect *their* values, but rather the values *in their environment*. On this understanding, the implicit prejudice is not a deep quality of the agent, and thus the card-carrying feminist cannot be said to be reprehensible for it. However, if the card-carrying feminist is *not* reprehensible, the obstructionist is committed to saying that one can be revision blameworthy for their vices without being reprehensible for them. Giving up the intuition that (revision) blameworthiness is a subset of reprehensibility constitutes the first horn of the dilemma.

Alternatively, one could argue that implicit prejudices are not explicitly held beliefs, however, they are still implicitly held beliefs that shape the kind of epistemic agent we

are. On this second understanding, the card-carrying feminist *is reprehensible* for implicit prejudice. But if they *are* reprehensible for implicit prejudice, despite it not reflecting their epistemic agency, it is unclear why cognitive defects that shape our implicit beliefs and motives (e.g. impaired vision) wouldn't equally count as deep qualities that shape the kind of epistemic agent we are. The second horn of the dilemma, according to Battaly, is accepting that epistemic agents are equally responsible for implicit bias (which is usually classified as a vice) as for certain cognitive defects. That is, obstructionism is committed to either excluding implicit bias as a vice or including cognitive defects like impaired vision as vices.

To recap: the card-carrying feminist isn't acquisition blameworthy but is revision blameworthy for her implicit bias. When we have to determine whether or not she is reprehensible for her implicit bias, neither route is satisfactory. We thus have a Responsibility Problem for vice epistemology.

3. Potential solutions

The Responsibility Problem leaves us with few options for proceeding with a non-voluntarist notion of responsibility for vice-epistemology. The first route is to embrace the first horn of the dilemma and (1) simply give up the intuition that blameworthiness is a subset of reprehensibility. This is broadly the route that Alessandra Tanesini (2020) advocates for. Tanesini points out that whilst Cassam does indeed implicitly commit to blameworthiness as a subset of reprehensibility, he only does so once. And on the other hand, he simultaneously allows for vice charging against ways of thinking that are 'out of character', e.g. occasional wishful thinking (Tanesini 2020: 856). This implies that agents can be held blameworthy for features that don't represent the kind of epistemic agent they are, and for which they are not reprehensible. Tanesini proposes that Cassam can respond to the Responsibility Problem by slightly changing his account and adopting a disjunctive account of responsibility, according to which agents can be blameworthy *or* reprehensible for their epistemic vices. A downside of this strategy is that it results in a lack of unity regarding the criteria for epistemic vices. However, as Tanesini argues, a Strawsonian backstory to our practice of holding people responsible can account for this disjunction. Tanesini distinguishes two motives for vice charging: we are either showing disdain for the kind of thinker they are, or we are blaming them for their vice. Accordingly, one can be held responsible for different kind of epistemic vices for different motives.

Tanesini's strategy is promising, even if it comes at the cost of giving up unity in criteria for epistemic vices. Nevertheless, I propose a second way out of the dilemma. Note that my strategy is compatible with Tanesini's proposal but doesn't require it. That is, proponents of my proposal can choose whether or not they want to give up the intuition that blameworthiness necessarily entails reprehensibility, or maintain unity in their criteria for epistemic vices. Additionally, my discussion of this alternative response to the dilemma brings out a different way of thinking about the characterization of implicit bias as an epistemic vice. The alternative strategy is to embrace the second horn of the dilemma and (2) accept that agents are equally responsible for implicit bias as they are for certain cognitive defects like impaired vision. According to Battaly, the problem here is that we would have to (2a) include cognitive defects like impaired vision as a vice or (2b) exclude implicit bias as a vice. I agree that the former (2a) seems problematic. However, (2b) excluding implicit bias as an (individual) vice might not be as problematic or theoretically inconsistent as Battaly portrays it to be.

I propose to embrace the second horn (2b to be precise) of the dilemma (2b to be precise) and exclude implicit bias as a vice. What we need in order to get around the Responsibility Problem whilst satisfying the intuition that epistemic agents should be held responsible for implicit biases, is an account of non-voluntarist responsibility that excludes implicit bias as a *vice* (such that cognitive defects that equally determine the kind of thinker we are not also counted as vices) – without letting the card-carrying feminist off the hook completely.

In fact, such a theory is readily available to us, but Battaly dismisses it (mistakenly, as I will argue) as being unable to deal with the card-carrying feminist case. Recall that Battaly states that the responsibility problem is not unique to Cassam's obstructionism, but rather, that all accounts of non-voluntarist responsibility for epistemic vices are vulnerable to this objection. She explicitly mentions Gary Watson's 'real self' view and Miranda Fricker's 'no-fault responsibility' as examples of other accounts that cannot escape the responsibility problem. Against Battaly's analysis, I argue that the Responsibility Problem does not apply to Fricker's theory. Fricker's notions of 'epistemic agent regret' and 'no-fault responsibility' do end up holding the card-carrying feminist equally responsible for her implicit bias as for cognitive defects like impaired vision (2). If we want to avoid reprehensibility for cognitive defects (2a), Fricker is committed to exclude implicit bias as a vice (2b). However – contrary to what Battaly claims – this is in fact theoretically compatible with Fricker's concept of testimonial injustice and other theoretical commitments.

Note that the Responsibility Problem is raised by pointing out difficulties in labelling implicit bias as an individual vice. My analysis is meant to illustrate that there are reasons to reconsider this label, which would likely resolve the Responsibility Problem. In fact, the classification of implicit bias as a vice might be all that Cassam has to give up in order to maintain his obstructionist definition of vice. The first horn of the dilemma (giving up the intuition that blameworthiness is a subset of reprehensibility, and thus giving up unity of criteria for vice) only emerges if we are committed to labelling implicit bias as a vice.

In what follows, I will first briefly summarize Fricker's version of non-voluntarist responsibility. I illustrate how Battaly's analysis of the card-carrying feminist case is flawed, as it fails to acknowledge that testimonial injustice and implicit prejudice are conceptually distinct. Testimonial injustice reflects badly on the kind of epistemic agent one is, but having implicit bias does not. Second, I argue that excluding implicit prejudice as a vice is not as problematic for Fricker's theoretical commitments as Battaly seems to assume. Third, contrary to popular belief, I argue that it isn't necessarily problematic to assign equal responsibilities for implicit prejudices as for cognitive defects like impaired vision.

4. Fricker's no-fault responsibility

Fricker's non-voluntarist notion of responsibility is rooted in the idea that the category of features for which we are responsible is broader than the category of features for which we are blameworthy. She builds on this notion when assessing the extent to which we are responsible for our implicit prejudice. Fricker uses a working definition of implicit prejudice as a 'dominant subset' of how Jules Holroyd (2012) conceptualizes implicit bias:

'An individual harbors an implicit bias against some stigmatized group (G), when she has automatic cognitive or affective associations between (her concept of) G and some negative property (P) or stereotypic trait (T), which are accessible and can be operative in influencing judgment and behaviour without the conscious awareness of the agent.' (Holroyd 2012: 275).

Implicit prejudice can then be understood as such associations that result from ‘*any motivated failure to properly gear one’s attitudes to the evidence*’ (Fricker 2016: 9). Fricker argues that agents can be held responsible for their implicit prejudices, even if they are non-culpable for acquiring them. This claim is supported by her consideration of cases wherein the emotion of *epistemic agent-regret*, the epistemic application of Bernard Williams’ *moral agent-regret*, seems appropriate (Fricker 2016: 35). Moral agent-regret is appropriate when the moral agent did something morally bad but did so blamelessly. On the epistemic equivalent we might think of someone who thinks badly but does so blamelessly. Epistemic agent regret is appropriate when someone makes a prejudiced judgement, but when the implicit prejudice is acquired through no fault of her own. For example, imagine the card-carrying feminist passively and innocently inherits gender prejudice, and has no irrational motivation for having this bias. In this case, the fault of motivated maladjustment to the evidence doesn’t originate in the epistemic agent, but from their environment, and ‘merely flows through her’ (Fricker 2016: 42). Fricker goes as far as to say this is a form of ‘environmental epistemic bad luck’ (Fricker 2016: 45), which prevents the agent from recognizing the epistemic significance of their flawed judgement. Consequently, they are not blameworthy for possessing the prejudice, but they should nevertheless ‘own up’ to resulting flawed judgements.

As previously mentioned, Battaly claims that this view is vulnerable to the Responsibility Problem. Her objection states that if agents ought to ‘own up’ to bad epistemic results from their implicit bias even they are not culpable for them, then the same should be said for bad result from cognitive defects (for which they are also not culpable). According to Battaly, Fricker’s no-fault view thus casts the net of responsibility too widely, including cognitive defects (Battaly 2019: 32). I argue instead that the no-fault responsibility view can be saved from this objection by showing that actually, it can embrace route 2b and exclude implicit prejudice as a vice. I argue that this is a possible and desirable move for Fricker, that does not undermine any of her explicit theoretical convictions.

One reason that could explain why Battaly thinks Fricker runs into problems here is due to her equating the vice of testimonial injustice with implicit bias. As I will illustrate, this is a mistake. Though the two concepts are linked, they are conceptually distinct. Having implicit prejudice doesn’t automatically imply testimonial injustice. The latter requires that the prejudice goes undetected or uncorrected by the hearer – Fricker herself says as much when she recaps some examples of testimonial injustice as instances whereby the hearers are ‘*failing to correct for identity prejudice in their testimonial sensibility*’ (Fricker 2007: 89, italics in original).¹

This conceptual distinction is also apparent in Fricker’s conceptualization of the virtue of testimonial justice. On Fricker’s account, the virtue of testimonial justice requires the hearer to *reliably neutralize* prejudice in their credibility appraisals or having unprejudiced credibility judgements (Fricker 2007: 92–93). Crucially is that the credibility judgement can be unprejudiced *naively*, i.e. when the hearer simply doesn’t hold any (relevant) prejudices, or in *corrective* form, whereby the hearer checks for and corrects their identity prejudices (Fricker 2007: 93). Since it is virtually impossible to not inherit any implicit biases, the virtue of testimonial justice typically looks like the latter. This is why Fricker puts so much emphasis on hearers practicing ‘reflexive critical social awareness’ (Fricker 2007: 91) as a way of detecting prejudices. Again, the reflective nature of the virtue of testimonial justice implies that the mere presence of implicit prejudice doesn’t necessarily prevent one from being testimonial just, as long as one is

¹This also fits the interpretation of testimonial injustice as a form of epistemic negligence, as suggested by e.g. Maitra 2010 and Riggs 2012.

able to reliably neutralize its impact on credibility judgements. On this reading testimonial injustice is a vice, but implicit prejudice need not be a vice per se.

A similar point can be derived from Fricker's discussion on how the notion of appropriate epistemic agent-regret for prejudiced thinking 1) opens up the possibility of being responsible yet not-culpable for their implicit prejudice and 2) shows how one can have epistemic obligations regarding 'owning up to' the implicit thinking (Fricker 2016). She gives an example of virtuous agents taking steps to correct for or prevent prejudice impact (e.g. implement anonymised review-processes to circumvent gender-prejudiced judgement). This again implies the possibility of a virtuous agent who possesses implicit prejudice – as long as they correct for them – and is compatible with excluding [some] implicit prejudices as a vice. After all, there is no reason to think we cannot be responsible for – or have epistemic agent-regret for – epistemic failings other than epistemic vices.

Note that this shifts the horn of the dilemma: Battaly mentions as one of the strengths of Fricker's view that it can account for the implicit bias as a vice, because we can hold people responsible for non-culpably acquired yet problematic epistemic features. She then objects to Fricker's view as (2a) throwing the net of responsibility too wide, including cognitive defects (Battaly 2019: 32). However, Fricker herself never explicitly refers to implicit prejudice as a vice. As I have argued, if anything, Fricker's view casts the net 'too narrowly' (in Battaly's words), excluding implicit bias as a vice.

Fricker's view is *compatible* with excluding implicit prejudice as a vice – but is this a desirable move? Battaly asserts that it is not, but this begs the question. Still, a perhaps less controversial option available to Fricker would be to hold that if implicit bias is a vice, it is nevertheless not an *individual vice*. This is still compatible with the premises she already holds. Recall that Fricker models her understanding of implicit prejudice after Holroyd's conception of implicit bias. In recent work, Holroyd (2020) has argued that implicit bias can best be understood not as an individual epistemic vice, but as a *collective vice*. This claim is based on empirical data that suggests that implicit biases are not predictive of individual epistemic judgements or behaviour, but that they are reliable indicators for epistemic/practical outcomes on the group level. On this understanding, implicit bias is, if you will, a contextual matter. This interpretation echoes Fricker's description of innocently inherited implicit prejudice as environmental epistemic bad luck. It should be noted that to my knowledge, Fricker has not explicitly agreed with the suggestion that implicit bias is a collective vice. However, given that she takes Holroyd's (2012) conception of implicit bias as starting point, this further claim is at least compatible with her theory.

5. Implications for embracing the second horn of the dilemma

As I have argued, there are two ways in which we can resolve the Responsibility Problem for Fricker's no-fault responsibility view: either by excluding implicit bias as an epistemic vice altogether, or by maintaining that the vice operates merely at a collective level. The implication of both routes is that the card-carrying feminist's implicit prejudice should be seen as a contextual feature, not as a feature of her epistemic agency, nor as reflecting her true self. The card-carrying feminist can however be blameworthy or otherwise reprehensible *for not trying to detect or correct for her prejudice*. Should she fail to try and correct for these prejudices, *that* fault would be located in her own agency. That is, she can be blameworthy or otherwise reprehensible for her testimonial injustice, even if she isn't reprehensible for the implicit prejudice on which the testimonial injustice is based. I.e. testimonial injustice can be labelled as an epistemic vice without labelling implicit prejudice as a vice.

I mentioned earlier that this strategy is compatible with Tanesini's route (1) of adapting Cassam's obstructionism according to a disjunctive account of responsibility

(Tanesini 2020). Nothing in what I've sketched out above prevents us from also adopting this move, so if we want to keep the strengths of her proposal (i.e. more deeply reflecting diverse motives behind vice charging), we can still do so. But for those who do not want to give up unity of criteria for epistemic vice; as illustrated above, my strategy doesn't require this. That is because on my reading of Fricker's account, the card-carrying feminist is not labelled as blameworthy for her implicit bias, but merely as no-fault responsible. Any potential blameworthiness or reprehensibility is aimed at the testimonial injustice. In other words, my analysis of the card-carrying feminist is compatible with the intuition that blameworthiness implies reprehensibility and does not require sacrificing unity in our definition of epistemic vice. However, *responsibility* does not imply blameworthiness. Responsibility and blameworthiness are often used interchangeably – which I think is also why Battaly assumes Fricker to label implicit bias as a vice just because she holds agents responsible for them. However, as Tanesini (2020) rightfully remarks: blameworthiness is again a subset of responsibility. Importantly, also responsibility and reprehensibility are conceptually distinct, and an agent can be responsible for features for which they are not reprehensible. After all, the notion of epistemic agent-regret is meant to capture exactly those cases whereby someone's agency is involved (i.e. they do something bad or think badly) without their actions reflecting badly on them. The fact that we can be responsible without being reprehensible is amplified in those cases where feelings of agent-regret are appropriate even if the reason for that regret is beyond the agent's control – and it seems that we find ourselves in those cases more often than we would like. I thus suggest that Fricker embraces the second horn of the dilemma and concedes (2) that we can be held equally responsible for implicit bias as for certain cognitive defects, and (2b) that implicit prejudice is not an (individual) vice. This resolves the Responsibility Problem for her view without giving up on any of her theoretical commitments.

Even if excluding implicit bias as an individual vice is compatible with the no-fault notion of responsibility, we must also ask whether this outcome is desirable for vice-epistemologists. There exists a widespread assumption in the vice epistemology literature that implicit bias is an individual vice. I suspect why most vice epistemologists (including Cassam and Battaly) insist on this characterization is because we all agree that aside from its knowledge-obstructing features, we think there is something wrong with reasoning from implicit bias, in such a way that it amounts to more than simply an epistemic error. It is one of these epistemic failings for which we want to hold epistemic agents responsible. But here is the thing; holding people responsible for their epistemic conduct doesn't only apply to epistemic vices. The no-fault notion of responsibility underlines how we can be responsible for features for which we are not reprehensible, without giving up the intuition that we are blameworthy or otherwise reprehensible for our epistemic vices.

If we can be held responsible for epistemic features without being reprehensible for them, this implies that we are similarly responsible for cognitive defects. Battaly presents this as a problem for Fricker's view. She suggests that it seems odd to hold agents equally responsible for impaired vision as for implicit prejudice, as the latter seems somehow more culpable. I argue instead that ascribing similar kind of responsibility to implicit prejudice and (certain) cognitive defects is actually fitting and will illustrate below how the no-fault responsibility applies to both epistemic features.

We often hold people responsible for erroneous judgements resulting from cognitive defects, but not in a way that finds them blameworthy or reprehensible for *having* the defect. Rather, one can be blameworthy or otherwise reprehensible for not correcting for these defects. I would argue that so understood, holding people equally responsible for implicit bias as for cognitive defects like impaired vision is actually an appropriate response if we use the no-fault notion of responsibility. Recall that the epistemic agent-regret of the

card-carrying feminist is not directed at the implicit prejudice itself but at judgements resulting from it. Similarly, we can imagine cases whereby someone's impaired vision causes epistemic harms, for which they are not culpable, yet no-fault responsible. They feel epistemic agent-regret for the bad outcomes, even in cases where they were not in a position to correct for their cognitive defects.

This parallel can be drawn even further when considering cases of *blameworthy* behaviour: We have established that one can rightfully be called epistemically vicious (i.e. blameworthy) for neglecting to correct their prejudiced thinking when they could have (e.g. in cases of testimonial injustice), even if they are not culpable for possessing the implicit prejudice in the first place. By the same token, one could be equally blameworthy for neglecting to correct for impaired vision when they could have (e.g. by wearing glasses), even if they are non-culpable for having impaired vision. If they are in a context wherein impaired vision carries foreseeable risk of epistemic or moral harms (e.g. when driving a car), their non-culpably acquired cognitive defects comes with special obligations, just like agents have special obligations to correct for non-culpably acquired implicit biases. Like with implicit prejudice, any blameworthiness or reprehensibility targets the negligence to comply with these obligations, not the cognitive defect itself. Additionally, the agent can be no-fault responsible for their cognitive defects without being reprehensible for having the cognitive defect. The fact that epistemic agents are ascribed the same kind of responsibility for their implicit bias as for their cognitive defects is thus not as counterintuitive as Battaly claims and does not undermine my claim that the no-fault responsibility view successfully avoids the Responsibility Problem.

Conceding that we are equally responsible for implicit bias as for some cognitive defects, and excluding implicit bias as a vice, absolves the Responsibility Problem. I want to emphasize that this move is available to other vice epistemologists too, once we properly separate the concepts of implicit bias and testimonial injustice. We can label testimonial injustice as a vice for which one is blameworthy or otherwise reprehensible, whilst assigning mere no-fault responsibility to implicit bias and excluding the latter as epistemic vice. This allows us to maintain the intuition that blameworthiness is a subset of reprehensibility and would allow Cassam to keep unity in his obstructionist definition of epistemic vice.

6. Conclusion

This paper engaged with Battaly's Responsibility Problem, which stated that non-voluntarist vice epistemologists are committed to either (1) counting some vices as blameworthy yet not reprehensible or (2) hold agents equally responsible for certain cognitive defects as for some vices. The second route would imply that (2a) cognitive defects are included as epistemic vices, or (2b) that implicit bias is excluded as epistemic vice. In applying Miranda Fricker's 'no-fault responsibility' view to the card-carrying feminist case, I have argued that Fricker can embrace route (2b) without giving up any theoretical commitments, and without having to endure problematic implications. I have shown that Battaly's analysis of the card-carrying feminist case is flawed, because it fails to acknowledge that the vice of testimonial injustice and implicit prejudice are conceptually distinct. In reconstructing the case, I have argued that excluding implicit prejudice as a vice is not as problematic as Battaly claimed, nor is assigning equal responsibilities for implicit prejudices as for cognitive defects like impaired vision. Being willing to (2b) exclude implicit bias as a vice seems a plausible way out of the Responsibility Problem for other vice epistemologists as well.

I thus suggest vice epistemologists to reconsider the classification of implicit bias as a vice, or – following Holroyd – at least reconsider the classification as individual vice. This isn't to downplay the significance of negative consequences resulting from implicit bias. Rather, my analysis is meant to clarify that in cases where we *do* find the agent blameworthy for negative epistemic outcomes resulting from implicit bias, the blameworthiness should target their *negligence of detecting and correcting* for their implicit biases. If vice charging is at least in part meant to express our disapproval or signal to others that this is behaviour should be avoided, it is important to direct those attitudes to the appropriate features.²

References

- Battaly H.** (2019). 'Vice Epistemology Has a Responsibility Problem.' *Philosophical Issues* 29(1), 24–36.
- Cassam Q.** (2019). *Vices of the Mind*. Oxford: Oxford University Press.
- Fricker M.** (2007). *Epistemic Injustice. Power and the Ethics of Knowing*. New York: Oxford University Press.
- Fricker M.** (2016). 'Fault and No-fault Responsibility for Implicit Prejudice. A Space for Epistemic Agent-regret.' In M. Fricker and M. Brady (eds), *The Epistemic Life of Groups: Essays in the Epistemology of Collectives*, pp. 33–50. Mind Occasional Series. Oxford: Oxford University Press.
- Holroyd J.** (2012). 'Responsibility for Implicit Bias.' *Journal of Social Philosophy* 43(3), 274–306.
- Holroyd J.** (2020). 'Implicit Bias and Epistemic Vice.' In I.J. Kidd, H. Battaly and Q. Cassam (eds), *Vice Epistemology*, pp. 126–147. London: Routledge.
- Maitra I.** (2010). 'The Nature of Epistemic Injustice.' *Philosophical Books* 5(4), 195–211.
- Riggs W.** (2012). 'Culpability for Epistemic Injustice: Deontic or Aretetic?' *Social Epistemology* 26(2), 149–62.
- Tanesini A.** (2020). 'Blaming the Intellectually Vicious: A Critical Discussion of Cassam's Account of Blameworthiness and Reprehensibility for Epistemic Vice.' *Ethic Theory and Moral Practice* 23(5), 851–9.

Carline Klijnman is a university lecturer and tutorial lecturer at Utrecht University and Eindhoven University of Technology. She will soon take up a position as an IRC Postdoctoral Research Fellow at University College Dublin. Her research interests include epistemic democracy, moral responsibility, social epistemology and epistemic injustice.

²I would like to thank Valeria Ottonelli, Jochen Briesen and audience members at the 2022 René Descartes Lectures 'Taking Responsibility' (University of Tilburg) for helpful comments and discussions on earlier versions of this paper.

Cite this article: Klijnman C.J.F. (2025). "Maintaining Reprehensibility for Epistemic Vice: Responsibility for Implicit Bias as Non-vicious Conduct." *Episteme* 1–10. <https://doi.org/10.1017/epi.2024.40>