

# Epistemic Duty and Implicit Bias

LINDSAY RETTLER AND BRADLEY RETTLER

forthcoming in *Epistemic Duties*

---

## 1 Introduction

Here's a riddle: A father and son are on their way to the zoo when they get into a car crash. They are rushed to the hospital and taken to separate operating rooms. A surgeon rushes into the boy's room, but upon looking at him, says, "I can't operate on this boy; he's my son!" How is this possible?!

Here's another: A mother and son are on their way to the zoo when they get into a car crash. They are rushed to the hospital and taken to separate operating rooms. A surgeon rushes into the boy's room, but upon looking at him, says, "I can't operate on this boy; he's my son!" How is this possible?!

The answer to both riddles is that the surgeon is the boy's other parent. The first of these is very common. We have never heard the second, despite the fact that it differs from the first in only one word—'mother' instead of 'father'. Why is this? Presumably it's that people who tell riddles think that in the second riddle, we would come up with the idea that the surgeon is the boy's father. Apparently it's much more difficult to think of a mother being a surgeon. But when asked, "Can a mother be a surgeon?", far more people would say "yes" than correctly answer the first riddle.

People's difficulty answering the riddle illustrates implicit bias. People explicitly say that mothers can be surgeons, and yet when all that it takes to answer a riddle is to consider the possibility that a mother is a surgeon, people have trouble doing so. This may be an example of gender bias—people don't think of women being surgeons. Or it may be that people have no intellectual problems with women becoming surgeons as long as those women don't have children. Of course, there are many kinds of implicit biases: race, gender, ethnicity, accent, ideology, size, age, and so on.<sup>1</sup>

Implicit biases are incredibly common. In fact, according to the Kirwan Institute for the Study of Race and Ethnicity, "Everyone possesses [implicit biases], even people with avowed commitments to impartiality such as judges."<sup>2</sup>

In this paper, we argue that people have an epistemic duty to eradicate at least

---

<sup>1</sup>One popular test to measure implicit bias is the Implicit Association Test: <https://implicit.harvard.edu/implicit/>. Some concerns with the test can be found in Azar (2008).

<sup>2</sup>See Kirwan Institute for Race and Ethnicity (2015). For more on the pervasiveness of implicit bias, see Jost et al. (2009).

some of the implicit biases they have.<sup>3</sup> We begin, in §2, by giving a few examples of implicit bias and noting a few important features. In §3, we clarify the notion of epistemic duty with which we're concerned. In §4, we offer the main argument: implicit biases prevent us from fulfilling our other epistemic duties, we ought to eradicate mental states that prevent us from fulfilling our other epistemic duties, so we ought to eradicate implicit biases. We conclude in §5 by responding to what we take to be the strongest objection: we only have duties to do things that are under our control, and our implicit biases aren't under our control, so we can't have duties to eradicate them. In our view, we have *indirect reflective control* over our implicit biases.

## 2 Implicit Bias

Let's start with two examples of implicit bias.

TED: At the age of 22, Will applied for a job working as a bank teller after finishing his bachelor's degree in Business Administration. He was eventually called in for an interview with the VP, Ted, which went very well, and Will was offered the job. He eagerly accepted. His career goal was to work his way up to a branch management position. He learned quickly, was proactive in taking on more responsibilities and attending training seminars, always came in on time and never missed work, and was overall an excellent employee. Every time a promotion came available, Will applied for the job. But promotions were slow in coming. Will was often passed over for jobs in favor of people who seemed less qualified. When asked why, Ted couldn't point to any particular feature of Will's resume or interview that threw up red flags, but Ted mentioned in passing that Will just didn't quite fit the profile of the role for which he was applying. But Ted was appalled at the suggestion that it had anything to do with the fact that Will is black.<sup>4</sup>

GINA: Gina is a recently tenured philosophy professor at a state school. Much of the teaching that she does is for a required Introduction to Ethics course. Rather than lecture, Gina likes to run the course as a kind of guided discussion, regularly throwing out provocative questions for the students to answer out loud. She has become frustrated that the women in the class tend to answer these questions with "feelings" language. She often reminds the women after they speak that in philosophy, reasons

---

<sup>3</sup>This may or may not entail that people also have a duty to *identify* the implicit biases they have. We don't take a stand on that. People may think that they *might* have an implicit bias, and thereby take steps to eradicate it; or they might simply do things that help eradicate the bias without being aware that they have it.

<sup>4</sup>This is a fictional story but very similar things happen all the time. See Bertrand and Mullainathan (2004).

trump emotions. Once a year, her teaching is evaluated by a peer and the department head. This year, in his evaluation, her peer writes that she mildly chastised a woman for starting a response with, “I feel that...”, but didn’t do so for a man who started his answer in the same way just a few minutes before. This surprised Gina, and so the next few classes she specifically paid attention to how everyone started their responses. She realized that, contrary to her impression, the men in the class used feelings language just as much as the women.<sup>5</sup>

Ted doesn’t explicitly *believe* that black people shouldn’t be branch managers. Gina doesn’t explicitly *believe* that women are more likely than men to use emotive language when discussing philosophical arguments. And yet both believe and act in ways that betray biases towards certain groups. These are *implicit* biases. There are a number of ways to carve out the difference between implicit and explicit biases. Perhaps implicit biases are the result of automatic processing, whereas explicit biases are the result of controlled processing. Perhaps implicit biases are unconscious, whereas explicit biases are conscious. We won’t take a stand on the correct analysis of implicit biases, or whether there is a unified analysis.<sup>6</sup> But we do conceive of an implicit bias as a mental state or process, rather than an output, such as a biased decision or action.<sup>7</sup> And one important question that we’ll return to later in section 4 is whether implicit biases are propositional attitudes or not.

If implicit biases are propositional attitudes, then the correct locution is “S’s implicit bias that  $p$ .”<sup>8</sup> And the propositions involved can be true or false, so implicit biases can be described as being true or false. It still would remain a question whether implicit biases are *sui generis* propositional attitudes or whether they can be subsumed under another propositional attitude, like belief.

But implicit biases as propositional attitudes is the minority view. The predominant view in both psychology and philosophy is that implicit biases are something like associations, presumably between concepts.<sup>9</sup> For example, according to Jules Holroyd (2012), “An individual harbors an implicit bias against some stigmatized group (G), when she has automatic cognitive or affective associations between (her

---

<sup>5</sup>This too is a fictional story. Schwitzgebel (2010) offers a similar example of a racially biased American philosophy professor, p532. See also Holroyd et al. (2017) p2.

<sup>6</sup>The interested reader should see Holroyd and Sweetman (2016).

<sup>7</sup>For more on this distinction, see Holroyd and Sweetman (2016) p81-82.

<sup>8</sup>See De Houwer (2014). Some advocate for the view that implicit biases are beliefs of some kind, or are based on beliefs, and thus have propositional structure: Frankish (2016), p26-29; Mandelbaum (2016); Hughes et al. (2011), and Mitchell et al. (2009). Schwitzgebel (2013) argues that implicit biases are “in-between” beliefs. For criticism, see Levy (2015), who argues that implicit biases are “patchy endorsements”, and see Madva (2016).

<sup>9</sup>See, e.g., Fazio (2007), Gawronski and Bodenhausen (2011), Gendler (2008b) p577, Gendler (2008a) p641-642, Leslie Ashburn-Nardo (2001), Holroyd (2012), Rydell and McConnell (2006), Saul (2013b) p244.

concept of) G and some negative property (P) or stereotypic trait (T)...” (p275). On this view, the correct locution is “S’s implicit association between *F*s and *G*s”. So, the propositional view has it that the implicit racial bias against Will is of the form, “S’s implicit bias that black people aren’t professional”, whereas the associative view puts it, “S’s implicit association between black people and unprofessionalism.” Associations between concepts or predicates aren’t truth-apt, so on this view implicit biases can’t be true or false. In fact, it’s difficult to see how associations like this can be evaluated.

What’s important for our purposes is whether we have an epistemic duty to eradicate implicit bias, and relatedly, whether we have the right kind of control over implicit bias. The argument we offer is compatible with multiple conceptions of implicit bias, but partly depends on whether implicit biases are propositional. We now turn our attention to epistemic duties.

### 3 Epistemic Duties

There are three important clarifications of the duty we’re arguing for. First, the duty is *epistemic*, as opposed to moral, or social, or pragmatic, etc.<sup>10</sup> Second, the duty is *robust*, as opposed to weak or attenuated in some sense. And third, the duty pertains to *eradicating* certain kinds of implicit bias, as opposed to not having such biases in the first place.

We likely have many duties regarding implicit bias. For example, perhaps we have a moral duty to not *act* on an implicit bias if doing so would harm another person. Maybe we have a social duty to call other people out on their implicit biases when we notice them. Perhaps we have a professional duty to proactively figure out if we have implicit biases, the having of which would compromise our ability to do our jobs well. These are not the duties we are interested in in this paper. The duty we’re interested in here is epistemic in nature. Epistemic duties are duties that we have with respect to the epistemic realm—belief, evidence, knowledge, etc. But that’s not enough to adequately distinguish them, because we might have other kinds of duties with respect to the epistemic realm, like moral duties. We *might* be morally obligated to, say, have certain beliefs or acquire certain evidence.

While we don’t have a definitive way to carve out what makes a duty epistemic, consider some paradigm examples of general epistemic duties that have been put forward. First, Chisholm (1989) suggests that we have a duty to believe that *p* if

---

<sup>10</sup>For an argument that epistemic duties aren’t just a special case of moral duties, see Haack (2001).

and only if  $p$  is true.<sup>11</sup> That is, we ought to: believe that  $p$  if and only if  $p$  is true. Chisholm thinks that it's plausible that this follows from our nature as intellectual beings. It is part of being a believer that we have this duty. But he also thinks that it is merely a *prima facie* duty, and often (and perhaps easily) overridden by other duties. One might have an epistemic duty to  $\phi$  but a moral duty not to  $\phi$ , and all things considered one ought not to  $\phi$ . Another suggestion is that we have the following epistemic duty: for every proposition of which I am not certain, I ought to look for more evidence regarding its truth.<sup>12</sup> Perhaps we have an epistemic duty to maximize true beliefs and minimize false ones.<sup>13</sup> Perhaps we have an epistemic duty to maximize justified beliefs. Perhaps we have an epistemic duty to maximize knowledge. Perhaps we have an epistemic duty to proportion our beliefs to our evidence.<sup>14</sup> Perhaps we have a duty to cultivate certain epistemic dispositions, like the disposition to believe a proposition when we know our evidence supports it, or as Sharon Ryan argues in this volume, the disposition to be open-minded. Perhaps there are many epistemic goals and many epistemic duties.<sup>15</sup>

Perhaps we even have epistemic duties to *act* in certain ways—maybe we have an epistemic duty to gather evidence. All these are examples of paradigm *epistemic* duties, as opposed to duties that are moral or social or pragmatic, etc.<sup>16</sup> One way to think about what unites these examples is that they are duties we have *qua* believers or knowers, etc.<sup>17</sup> But one could be forgiven for thinking this is circular, and so not an illuminating analysis of what makes a duty epistemic. Hopefully we have said enough to distinguish epistemic duties from other duties. Whatever our epistemic duties are, implicit biases get in the way of fulfilling those duties. So, our duty to eradicate them is connected with our other epistemic duties. (This argument is the topic of the next section.)

Our second clarification is that the epistemic duty to eradicate implicit bias is what we'll call a *robust* obligation. Robust obligations are such that their violation makes one *blameworthy*. If we fail to eradicate implicit bias, then we can be legitimately blamed for this failure, as long as there are no exonerating or excusing conditions in

---

<sup>11</sup>This occurs on p14. Roderick Firth (1978) p224 also discusses a duty to believe propositions if and only if they are true, but he does not ultimately endorse it.

<sup>12</sup>See, e.g. Hall and Johnson (1998) and Corlett (2003) p68.

<sup>13</sup>Truth seems to be a popular epistemic *goal*. See David (2001) and Hess (2010). For an argument against the view, see Kvanvig (2005).

<sup>14</sup>See Stapleford (2012).

<sup>15</sup>See DePaul (2001) and Kvanvig (2005).

<sup>16</sup>Of course, for some duties, there is likely some overlap between the epistemic dimension of the duty and other dimensions. For example, Jennifer Lackey (forthcoming) argues that the source of our duty to object to things that people say has both an epistemic dimension and a moral one.

<sup>17</sup>See, e.g. Feldman (1988): "The peculiarly epistemic judgment concerns not practical merits but rather the propriety of a disinterested believer in [the] situation" (p236).

place. The idea is that we are responsible for eradicating implicit biases.

This contrasts with a view according to which violation of the epistemic duty makes the agent in question worthy of, at most, criticism or negative evaluation. On this sort of view our epistemic duty to eradicate implicit bias is what we might call a *weak* obligation.<sup>18</sup> In other words, it's merely a standard for evaluation. Accordingly, if we fail to eradicate implicit bias, then we can be criticized or judged to have failed the standard, but not blamed.

Moral duties are paradigm examples of robust obligations: if you violate the duty not to wantonly harm an innocent person, then you can be legitimately blamed for this failure. While there are many different accounts of the nature of moral blame, it's fairly agreed upon that moral blame has “a characteristic depth, force, or sting” that goes beyond merely evaluative grading against a moral standard.<sup>19</sup> This force can be explained in many ways, but we cast our lot with those who hold that to morally blame someone is to place a normative demand on her.<sup>20</sup> Moral blame involves a demand for justification on the basis of moral standards we expect reasonable people to accept.<sup>21</sup> The demand is that the agent acknowledge her moral failure given the moral reasons she had. As it is with moral blame, so also with epistemic blame, i.e. blame for violating epistemic duties: blaming someone for violating a robust epistemic duty involves demanding that the agent acknowledge her epistemic failure given the epistemic reasons she had.<sup>22</sup>

In this paper, we do not offer arguments for the existence of robust epistemic duties. We assume that there are such duties, i.e. that at least some of what philosophers call “epistemic duties” are not merely standards of epistemic evaluation. Accordingly, the epistemic duty to eradicate implicit bias is not merely a standard of epistemic evaluation. It's a robust duty, such that failure to fulfill it merits blame.

It's important to note that we're not claiming that all epistemic duties are robust. Whether an epistemic duty is robust—i.e. whether an agent is blameworthy for failing to fulfill it—depends on whether the agent has the right kind of control. Weak duties, i.e. merely evaluative standards, do not require that an agent has control. Suppose Maya shows up for an open college basketball tryout and stands in the group wanting

---

<sup>18</sup>Stapleford (2012) distinguishes “oughts” from “duties” and carves out what he calls a “deflationary sense of ‘ought’ ” (p150, and p172, fn 8). What we're referring to as a “weak duty” or “weak obligation” is what Stapleford refers to as a “deflationary ought”.

<sup>19</sup>See Hieronymi (2004), p116-117. See also Scanlon (1998) p269, Smith (2008) p369, and Smith (2013) p29, Wallace (1994) p80-81, and Wolf (1990) p40-41, 64).

<sup>20</sup>See Hieronymi (2004) p117, McKenna (1998), Strawson (1974) p6-7, 14-16, and Wallace (1994), Wallace (2007), Wallace (2010), and Wallace (2011).

<sup>21</sup>See Smith (2008) p369.

<sup>22</sup>See Rettler (2018b), p2208-2210 for more detailed discussion of the normative demand inherent in epistemic blame.

to play center. The coach looks everyone over, and her 5'4" height is noticeable. He says, "If you want to play center, you should be taller." He has evaluated her as falling short of the height norm for NCAA centers. This is a legitimate judgment, despite the fact that it is something over which she has no control. However, if Maya were to deserve blame for failing the NCAA height standard—i.e. if the standard was a robust duty—then she would have to have control over her height. Similarly, in order for agents to be subject to the epistemic duty to eradicate implicit bias, they must have control over whether they eradicate such bias.

Our third clarification relates to the specific epistemic duty we defend: that of eradicating implicit bias. The duty in question is to eradicate implicit biases *we already have*. It does not address the question of whether we should have acquired them in the first place. It may well be the case (although we suspect it is not) that we have a duty, epistemic or otherwise, to not acquire implicit biases. But that's not what we're concerned with in this paper.<sup>23</sup> We now turn our attention to the argument.

## 4 An Epistemic Duty to Eradicate Implicit Bias

Our argument proceeds by first noting that implicit biases have a negative epistemic impact. For example, in TED, Ted has formed the belief that Will is unqualified for the job for which he is applying. Ted's implicit bias has caused him to form this false and unjustified belief about Will (and likely other black people). In GINA, Gina forms the belief that her women students are using "feelings language" more than her men students, which is false and unjustified.

Several philosophers have discussed what are sometimes referred to as the "epistemic costs" of implicit bias. Jennifer Saul (2013b) details quite a few. For example, "shooter bias" can negatively influence our perception (245-246). Implicit biases lead to inaccurate judgments about the best job candidate, which papers deserve to be published, and who is a credible witness (246, 248). And in philosophy they can lead to inaccurate philosophical beliefs, given that they influence which arguments we accept and reject (247). Implicit bias influences which testimony we accept and reject, as well, which Saul points out can thus influence our evidence base for various beliefs (247-249). Uwe Peters (2019) discusses the epistemic costs of what he calls "ideological bias"—associations between various ideologies and positive or negative properties—and argues that in combination with other relevant factors, ideological bias threatens "reliable belief formation and comprehensive knowledge acquisition"

---

<sup>23</sup>The reason we're not concerned with it is that we don't think we have the right kind of control over whether we acquire implicit biases. So, we don't want to defend this duty.

in the field of philosophy (395).

This suggests that implicit biases prevent us from fulfilling our other epistemic duties — things like believing what’s true, believing what you’re justified in believing, not believing what you’re not justified in believing. They cause us to have false beliefs and unjustified attitudes, they block knowledge and understanding, they cause our credences to not match our evidence. And we ought to eradicate mental states that prevent us from fulfilling our other epistemic duties. So, since implicit biases prevent us from fulfilling our other epistemic duties, we ought to eradicate implicit biases.

Consider a parallel argument that establishes a *moral* duty to eradicate implicit biases: Implicit biases have a negative causal impact on our morality. We have moral duties regarding others’ well-being, our relationships with people, and to cultivate virtues like justice, compassion, empathy, kindness, and compassion. But having implicit biases causes agents to violate these moral duties. For example, Ted consistently passes Will over for jobs for which Will is better qualified than the person who gets it; that’s unjust. Gina might well grade women in her class more harshly than she grades men, since she implicitly associates women with emotive language and men with rational language. The general argument is that (i) we ought to fulfill our moral duties, (ii) having implicit biases prevents us from fulfilling our moral duties, so (iii) we ought not to harbor implicit biases. The conclusion, that is, is that we (morally) ought to eradicate our implicit biases.

The moral and epistemic arguments have the same structure. Both rely on the connection between implicit biases and other things—moral or epistemic duties.<sup>24</sup> It is in virtue of this connection that we have a duty to eradicate our implicit biases. This way of arguing for the epistemic duty to eradicate implicit bias is, one might think, instrumental or relational or extrinsic: it’s bad to have implicit biases because they cause other bad things. This certainly seems like a reasonable line of argument.

But there’s another potential line of argument; it might also be true that implicit biases are epistemically bad in and of themselves, i.e. intrinsically.<sup>25</sup> This may depend on whether implicit biases are propositional. If they are—e.g. if they turn out to be beliefs—then they can be evaluated as true or false, justified or unjustified. In that case, implicit biases are epistemically bad when they are false or unjustified, in addition to being epistemically bad when they cause other false attitudes.

If, on the other hand, implicit biases are non-propositional, then they can’t be

---

<sup>24</sup>The argument could also be filled out in terms of moral and epistemic *values*. We could say that we ought to pursue things of moral/epistemic value, implicit biases prevent us from realizing things of moral/epistemic value, therefore we ought to eradicate our implicit biases. We encourage the reader to use whichever formulation is most persuasive.

<sup>25</sup>See Kelly and Roedder (2008) for reasons why (in particular, racial) implicit bias is intrinsically *morally* bad.



evaluated with respect to truth or justification. However, we might still evaluate them as epistemically bad *pairings*. In “On being good at seeming smart”, Eric Schwitzgebel notes that some associations with concept of intelligence seem “good”, for example, poise, confidence, giving a moderate amount of detail but not too much, providing some frame and jargon, etc., but others are not: “whiteness, maleness, a certain physical bearing, a certain intellect.”<sup>26</sup> Other pairings that seem epistemically bad (even if not false), include associating black people with violence, women with timidity, children with insignificance, people of certain regions or religions with lack of intelligence. We can call these pairings “epistemically inapt”. Perhaps, as Daniel Kelly and Erica Roedder suggest, we can also evaluate implicit biases with respect to rationality depending on whether the association between concepts “accurately reflects a correlation or statistical regularity that holds among those referents of the concepts” (p529).

So, at the very least we have a duty to eradicate implicit biases because of their negative epistemic impact. But perhaps we have an additional motivation for the duty, if implicit biases themselves are either false, irrational, or in some other way inapt. This parallels what we might say about why we ought not to have a false belief (if one thinks we have such a duty): because it’s inaccurate—it fails to represent the world correctly—but also because it causes other epistemically bad attitudes.

But it seems that not all implicit biases have extrinsic negative epistemic impact as a matter of necessity. As Jennifer Saul (2013a) notes, “one might unconsciously associate groups with different flavors of ice cream without this having any negative effects” (p40, fn 4). Or the impact of the implicit bias might be what we could call “epistemically neutral”. For example, implicit biases can sometimes cause agents to confabulate explanations of our decisions and actions, which might well be considered a negative epistemic impact. However, Ema Sullivan-Bissett (2014) argues such confabulatory explanations might sometimes be epistemically innocent. In these sorts of cases, the implicit bias does not have a negative, but rather, a neutral epistemic impact. Additionally, an implicit bias might have positive epistemic impact. Consider an example from Uwe Peters: “the implicit ... negative evaluation and exclusion of some political view-points (e.g., extreme right-wing positions) and their advocates is arguable epistemically ... beneficial for philosophy” (2019, p395). Accordingly, we do not claim that agents have an epistemic duty to eradicate implicit biases that are apt *and* have no negative epistemic impact.

Some implicit biases are epistemically worse than others—namely those with *greater* negative epistemic impact. Therefore, the duty to eradicate these biases is stronger—i.e. more important—than the duty to eradicate other less epistemically

---

<sup>26</sup>See Schwitzgebel (2010) and a discussion in Saul (2013a), p53-54).

problematic biases. Which biases are worse than others is relative to the agent in question, and depends on her other beliefs, what she attends to, her professional sphere, her personal sphere, etc. For example, as professional philosophers it is epistemically worse for us to hold a bias against women *in philosophy* than it is for a person working the front desk at a hotel, because we hold many more beliefs that are likely to interact with the bias against women in philosophy (e.g. an association pairing women with being less capable of rational argument). So an implicit bias against women in philosophy will lead to more epistemic error in a philosopher than in a person working the front desk at a hotel. Thus the professional philosopher has a stronger epistemic duty to eradicate implicit bias against women in philosophy, than the front desk clerk.<sup>27</sup>

We'll conclude this section by noting that it might be the case that people have an epistemic duty to eradicate at least *some* implicit biases that actually have no negative epistemic impact. Here's why. Most of us have the sense that we shouldn't be able to avoid fulfilling duties purely as a matter of luck. For example, we have a duty not to drive while drunk. If someone is driving drunk and is lucky enough to not encounter another person on their whole trip, we do not say that they aren't blameworthy for their actions. They still violated a duty; they just got lucky. If there had been a person, they may well have killed them. And in plenty of nearby possible worlds, they do just that.

Similarly, we shouldn't be able to avoid epistemic duties purely as a matter of luck. If someone has an implicit bias against black people but never meets a black person and so doesn't form false and unjustified beliefs about that person, we do not say that they aren't blameworthy for having that bias. They still violated an epistemic duty. If they were to meet a black person, they'd form false and unjustified beliefs about that person. And in plenty of nearby worlds, they do just that. So maybe an implicit bias's having an actual negative epistemic impact isn't *necessary* for us to have a duty to eradicate it.

---

<sup>27</sup>Note that this is not a question of a professional obligation. Which professional duties a person has are relative to a person's profession, and agents can have professional duties to believe certain things. For example, a surgeon has a professional duty to believe correctly about the best surgical procedures and devices. Similarly, philosophers have a professional duty not to discriminate against women in philosophy, which is a duty which professional chefs do not have. But we're not talking about professional duties here; we're talking about biases that are worse for a philosopher to have given their greater potential for negative epistemic impact.

## 5 Control over Implicit Bias

We've argued that we have an epistemic duty to eradicate at least some implicit biases. But consider the following objection: if it turns out that we don't have control over them, then we *can't* be subject to a duty to eradicate them—even if it's true that having implicit biases has negative epistemic impact. A major part of defending the epistemic duty to eradicate implicit bias is to respond to this objection—to show that we have the right kind of control over implicit biases, such that we can fulfill this duty. Here's a formalization of what we'll call *the Control Challenge*:

1. If agents have an epistemic duty to eradicate implicit bias, then they have control over whether they eradicate implicit bias.
2. It is not the case that agents have control over whether they eradicate implicit bias.
3. Therefore, agents do not have an epistemic duty to eradicate implicit bias.

We grant (1); we think that having a duty to do something requires that the agent have some kind of control over that thing. As we explained in §4, the kind of duty we're concerned with is such that failure to fulfill it can engender legitimate blame. The sort of control that satisfies the control condition for robust duty thus has at least two features: (a) it enables the agent to fulfill or not fulfill the duty, and (b) it enables the agent to respond to blame were she to fail the duty.<sup>28</sup>

That leaves us with the task of rejecting (2). (2) seems initially plausible because implicit biases seem like the kind of thing we don't have control over. As mentioned in §2, if implicit biases are the result of automatic processing, then they occur regardless of whether the agent is attending to them. Moreover, on some understandings of implicit bias, people are not even aware that they have them. And if they're not aware that they have them, how could they control them? So Jennifer Saul (2013a) argues, “we should abandon the view that all biases against stigmatised groups are *blameworthy* ... [because a] person should not be blamed for an implicit bias that they are completely unaware of, which results solely from the fact that they live in a sexist culture” (p55, emphasis in original). Lack of awareness suggests (though does not entail) a lack of control, which in turn rules out legitimate blame.

---

<sup>28</sup>There may well be attenuated forms of blame that do not require control, such as blaming the weather for ruining one's picnic, but again, that's not the kind of blame involved in violating the epistemic duty to eradicate implicit bias.

Given our lack of awareness of implicit biases, it seems that introspection doesn't help us correct for them.<sup>29</sup> For example, reflecting on whether you have an implicit bias that associates women with emotionalism or a bias that associates black men with unprofessionalism will not necessarily reveal to you whether you have that bias. Additionally, it seems that implicit biases are not directly responsive to reflection.<sup>30</sup> Even if you come to realize that you have an implicit bias associating women with emotionalism and you explicitly believe that women are not emotional, your implicit bias does not automatically come into alignment with your explicit belief just by reflecting on the matter.<sup>31</sup> This resistance to reflection makes implicit biases particularly insidious. Another reason why implicit biases seem out of our control is that they seem to reoccur despite intentional efforts to eradicate them.<sup>32</sup>

We grant all of these points. But we think that what these considerations show is that implicit biases are not under our *direct* control. Specifically, we cannot form an intention regarding our implicit biases and then directly or immediately carry out that intention.<sup>33</sup> However, these considerations do not show that implicit biases are not under our *indirect* control. And we don't think that direct control is necessary for being subject to the duty to eradicate implicit bias.

We have many moral obligations regarding things over which we lack direct control: the duty to keep our promises, to be a good friend, to take care of the environment, to refrain from wantonly harming others, and so forth. None of these duties can be fulfilled directly.<sup>34</sup> Instead they are fulfilled *indirectly* by carrying out intentions to do other things. For example, for Duncan to carry out the intention to be a good friend, he must carry out intentions to spend time with his friends, to listen to them, to help them when they're in need, to say kind things to them...So,

---

<sup>29</sup>“Neither introspection nor honest self-report are reliable guides to the presence of such mental states, and one may harbor implicit biases that are diametrically opposed to one's explicitly stated and consciously avowed attitudes” Kelly and Roedder (2008) p532). See also Holroyd (2012), p275. See Ballantyne (2015), p150 for helpful discussion of the relation between bias and introspection.

<sup>30</sup>See Holroyd (2012), p275.

<sup>31</sup>See Gregg et al. (2006), and Levy (2016), p20 for discussion.

<sup>32</sup>See Huebner (2009), Follenfant and Ric (2010), Macrae et al. (1994), and Holroyd (2012) p283 for more discussion of this “rebound effect”.

<sup>33</sup>As Neil Levy (2016) puts it, we lack control that is “exercised in the service of an explicit intention (*to make it the case that such and such*)” (p8-9, emphasis in the original).

<sup>34</sup>For additional examples of things for which we hold agents responsible but over which we lack direct control, see Holroyd (2012) p284-285, Rettler (2018b) p2212-2213, and Smith (2005). Not all of their examples are examples of duties, but rather responsibility more generally. For example, Holroyd points out that we hold individuals responsible for cultivating various virtues though we only have indirect control over such cultivation. It's unclear whether we have a *duty* to cultivate virtue. But regardless, the same considerations regarding control apply: we are held responsible for failing to cultivate virtues even though we only have indirect control over doing so; we have a duty to cultivate virtues even though we only have indirect control over doing so.

it seems that with respect to moral duties, while direct control is *sufficient* to satisfy the control condition for being subject to a duty, it is not necessary. Instead agents can be subject to a duty even if they only have indirect control over fulfilling it.

Now consider epistemic duties. If we are subject to robust duties regarding our *beliefs*—i.e. if failure to comply with various doxastic duties makes agents blameworthy—then we must have control over our beliefs. But as William Alston has famously argued, we lack *direct* control over our beliefs—we cannot believe that *p* as the direct result of an intention to believe that *p*.<sup>35</sup> However, like moral duties, there is no reason to think that *direct* control is required for doxastic duties, so long as we have the right kind of *indirect* control over our beliefs.

Some have argued that we have a different kind of direct control over our beliefs that we might call reason-based, rather than intention-based. Instead of carrying out intentions regarding our beliefs, we control our beliefs by believing directly in response to various reasons and evidence.<sup>36</sup> It might seem that we can appeal to this “reason-responsiveness” as a form of direct control that satisfies the control condition for epistemic duties regarding belief. However, reason-responsiveness is not the right sort of control to appeal to in order to help explain why we are blameworthy for failing to comply with epistemic duties, because the kind of control that grounds blame must be *reflective*. And reason-responsiveness is not reflective.<sup>37</sup> (More on this below.)

In any case, appeal to reason-responsiveness seems even less plausible with respect to implicit biases, given that they are not responsive to reasons in the way that explicit beliefs are.<sup>38</sup> As noted above, our implicit biases do *not* necessarily reflect our rational judgments, and in fact often conflict with them. If Gina were to consider her reasons for believing that women are more likely to use feelings language than men, and conclude on the strength of the evidence that this claim is false, such a conclusion would not necessarily overturn her implicit association of women with emotionalism.

However, though we lack direct intention-based control over our beliefs, and though reason-responsiveness does not satisfy the control condition for legitimate blame, one of us has argued elsewhere that we do have a form of *indirect* control over our beliefs that *does* satisfy the control condition: *indirect reflective control*. An

<sup>35</sup>Alston (1988) p263-277, and see Rettler (2018b) p2212 for discussion of the argument.

<sup>36</sup>For various versions of this kind of view, see Ryan (2003) and Steup (2008), Steup (2011), Steup (2012), both of whom refer to their views as “doxastic compatibilism.” See also Hieronymi (2006) and Hieronymi (2008) p362-363; Shah (2002) p443; Smith (2000) p40-46 and Smith (2005) p236-271; and Weatherston (2008) p546).

<sup>37</sup>Or at least, that’s what one of us has argued and the other believes on the strength of that argument. See Rettler (2018b) p2213-2217.

<sup>38</sup>Even those who hold the view that implicit biases are beliefs do not advocate that they are as responsive to reasons as explicit beliefs. They are at best, minimally responsive. See Mandelbaum (2016) p645; and Kelly and Roedder (2008) pp528-529; and Holroyd (2012) p295.

agent has indirect reflective control over whether she believes that  $p$  if and only if she can actively engage in critical reflection that causally influences whether or not she holds the belief that  $p$ .<sup>39</sup> The reflection involved is *critical* because it is second-order reasoning involving conceptualization of reasons as such, and conceptualization of support relations between propositions and between our attitudes toward those propositions. This contrasts with what is sometimes called “merely first-order reasoning”. And the reflection involved is *active* in that an agent forms and executes intentions to do the mental actions that compose reflection. Engaging in reflection makes a causal difference to whether we have various beliefs, thus giving us indirect influence-based control over those beliefs. For example, we can ask ourselves various questions and try to answer them: “what are my reasons for believing  $p$ ?”, “are these good reasons?”, “is there any relevant evidence I haven’t considered?”, and so forth. We can direct our attention to various reasons, try to remember relevant data, reconstruct arguments, etc. All of these mental actions are part of reflection, and doing them causally influences what we believe (Rettler 2018b, p2217). This kind of control is both necessary and sufficient to satisfy the control condition for legitimate blame. It enables an agent to respond to the demand inherent in blame, which is a demand to acknowledge one’s failure. Without the capacity to actively engage in reflection that causally influences one’s beliefs, an agent would not be able to understand or respond to blame (Rettler 2018b, p2216).

We think that this kind of control is something we also have over many of our implicit biases—indeed, over all the implicit biases we can be blamed for failing to eradicate. However, the relationship between reflection and implicit bias is not as tightly linked as the relationship between reflection and belief; our control over implicit bias is even more indirect than our control over belief. With respect to believing that  $p$ , reflecting on the evidence for  $p$  influences whether we believe that  $p$ . But as we have seen, this is not necessarily the case with implicit bias: reflecting on whether women are more emotional than men will not necessarily overturn Gina’s implicit bias associating women with emotionalism.

However, some empirical research indicates that reflection on arguments suitably related to the group that one is implicitly biased against can have an impact on whether one harbors the bias. Briñol et al. (2009) conducted a series of experiments designed to test the impact of argument evaluation on implicit biases (what they call ‘automatic evaluations’), and they found that reflection on strong arguments does indeed make a difference to implicit biases.<sup>40</sup> In one particular experiment, differing participants were offered either strong or weak arguments in favor of integrating

---

<sup>39</sup>See Rettler (2018b) p2217 and Rettler (2018a) p103.

<sup>40</sup>See p291-297. See also Mandelbaum (2016) p640-641 for discussion.

more African-American professors into a university. Additionally, some were given additional prompts to encourage more thoughtful reflection on the arguments. When tested for implicit bias against black people, those given stronger arguments measured as having less implicit bias than those given weaker arguments. Moreover, the difference between the strong and weak argument groups was even greater between the groups prompted to think more carefully about arguments. These findings provide reason to think that reflection on strong arguments that cast the group one is biased against in a positive light make a difference to whether we have the biases. Thus, while Gina might not be able to eradicate her implicit bias by reflecting on reasons to not have it, it would help eradicate it if she were to reflect on (for example) strong arguments in favor of trying hard to recruit more women as philosophy majors.

Additionally, there seems to be evidence that implicit biases are causally related to various explicit beliefs and values. Insofar as reflection can causally influence these beliefs and values, reflection can causally influence whether we harbor implicit biases. Jules Holroyd argues that the manifestation of implicit bias is influenced by our explicit beliefs and values (2012, p288). For example, according to Holroyd's interpretation of Devine et al. (2002), "Individuals who are highly committed to responding without prejudice, for its own sake (and not rather or also for reasons of social pressure or norms), manifest significantly less negative race bias across a range of texts for implicit biases" (p288).<sup>41</sup> Though this research is concerned with control over whether we *manifest* our implicit biases, as opposed to control over whether we *have* implicit biases, the same idea applies to both having and manifesting bias. There are certain beliefs and values, the having of which influences whether we continue to have an implicit bias. Voluntarily engaging in reflection makes a difference to our beliefs about values, as well as to our values themselves. And insofar as engaging in reflection has this indirect influence on beliefs and values that are causally related to implicit bias, it has an indirect influence on implicit bias.

Consider an example of this influence. Jennifer Saul (2013a) argues that "The view that traits are fixed makes one more prone to stereotype endorsement and to implicit bias. If intellectual ability is viewed as a more complicated set of abilities and skills, both of these problematic phenomena are reduced" (p53).<sup>42</sup> Voluntarily engaging in reflection on the proposition that traits are fixed causally influences whether one believes this proposition, which in turn causally influences whether one has various

---

<sup>41</sup>See also Levy (2016) p. 20 for discussion.

<sup>42</sup>Saul even explicitly encourages philosophers to reflect on this point "...because it seems to me that philosophers are very prone to claims about "who's smart" and "who's stupid" " (p. 53). Such encouragement supports the view that reflection is something we can actively engage in, and that such reflection can make a difference to whether agents harbor implicit bias against women and against women in philosophy.

implicit biases.

The above considerations involve reflection—on arguments, on the content of various beliefs, and on values. This reflection makes a difference to whether we harbor implicit bias. But empirical research also shows that there are various other actions we can take that make a difference to whether we harbor implicit bias—various techniques that have been proven to help individuals rid themselves of their implicit biases.

One such technique is interacting with other social groups; the more people spend time with members of groups other than their own, the less biased they appear to be.<sup>43</sup> Additionally, long term exposure to individuals that counter stereotypes can help remove implicit bias.<sup>44</sup> Another is “approach training”. People are asked to push a button when they see a stereotype depicted on a television screen. After spending some time doing this, they measure as having less implicit bias than beforehand.<sup>45</sup> Another technique is showing pictures of admired people in a group that the subject is biased against and disliked people in a group that the subject is biased towards.<sup>46</sup> Yet another technique is to have light-skinned people interact in virtual reality as dark-skinned virtual bodies.<sup>47</sup> Finally, diversity education lowers implicit bias.<sup>48</sup>

When discussing implicit bias against women in philosophy, Jennifer Saul identifies the existence of stereotypes as part of the cause of implicit biases, and she argues that anything done to break down those stereotypes helps “alleviate the problem” (2013, p. 50). There are many ways to break down stereotypes. For example, Ted could hire Will, praise Will’s work to others, make sure the photos in the bank and on its website feature black bankers, and things like that.

Empirical research shows that doing these sorts of things influences whether one harbors implicit bias. We might conclude as Holroyd (2012) does, that we have what she calls “long-range” control over many of our implicit biases (p286-288).<sup>49</sup> We have this long range control over whether we eradicate implicit bias insofar as we can over time carry out intentions to do various things which then results in eradication of implicit bias.

But notice that long range control is a form of intention-based control: an agent

---

<sup>43</sup>See, e.g., Amir (1969), Cook (1985), Brewer and Miller (1988).

<sup>44</sup>See Dasgupta (2013), and Levy (2016) p. 6 for discussion.

<sup>45</sup>See, e.g., Kawakami et al. (2007), Phills et al. (2011a), Phills et al. (2011b), and Madva (2017).

<sup>46</sup>As in Nilanjana Dasgupta (2001).

<sup>47</sup>See Peck et al. (2013).

<sup>48</sup>See Rudman et al. (2001).

<sup>49</sup>Though Holroyd thinks the evidence more strongly supports the idea that we have long range control over *manifestation* of implicit biases, rather the having of such biases. See also, Alston (1988), pp274-277 for a discussion of what he calls “indirect long-range voluntary control” over our *beliefs*.



controls whether she does  $X$  by carrying out an intention to  $Y$ , which then over time results in her doing  $X$ . For example: an agent controls whether she harbors implicit bias  $B$ , by carrying out an intention to engage in approach training which over time results in eradicating  $B$ . But, we argue, long-range control by itself is not sufficient to satisfy the control condition for robust duties. If one is subject to a robust duty, then one can be legitimately blamed for failing the duty. But such blame is only legitimate if one has the kind of control that enables one to *both* satisfy the duty *and* to understand and respond to blame in case one fails the duty. To be able to respond to blame, agents must be able to acknowledge their failure, which requires the capacity for reflection. Suppose Gina fails to eradicate her implicit bias against women. She can't be blamed for failing to do so on account of the fact that she could have engaged in approach training, unless she can also recognize as a reason to engage in approach training that it will help eradicate her implicit bias. Without the ability to grasp this connection between approach training and eradicating implicit bias, Gina can't be held responsible for failing to eradicate her implicit bias via approach training. And to make this connection, Gina must have the capacity for reflection. In other words, she must be capable of engaging in reflection on whether to do approach training as a method to eradicate her implicit bias. Note that she doesn't have to *in fact* engage in such reflection; she just has to be *capable* of doing so. If she is not, then she can't be blamed for her implicit bias given her failure to engage in approach training. And the same is true for any actions or activities that make a difference to whether she harbors implicit bias.

Note also that Gina doesn't need to be aware of approach training as a method for eradicating her implicit bias. To be held responsible for failing her duty to eradicate her bias, it's enough that she *could* become aware of the various methods that would enable her to satisfy her duty. There are a number of things that it's plausible that agents must be able to be aware of in order to be subject to the epistemic duty to eradicate implicit bias: for example, that one has an implicit bias or that it's likely or possible that one does, that one has an epistemic duty to eradicate implicit bias, that implicit biases have negative epistemic impact, that there are various methods one can implement to help eradicate one's implicit bias, and so forth. To be able to become aware of these things, there must be evidence available to the agent, and she must be able to recognize and evaluate it. To the extent that an agent cannot become aware of these things—whether because the evidence out there is very difficult to acquire or because the agent has a hard time recognizing or evaluating it—the agent is not as blameworthy for failing to eradicate her implicit bias.<sup>50</sup>

---

<sup>50</sup>Alternatively, one might hold that if an agent is incapable of the requisite awareness, she is to that extent not subject to the duty.

To summarize: there are two ways in which we have indirect reflective control over whether we harbor implicit biases. One is that we can actively engage in reflection on particular arguments, as well as various beliefs and values that we hold, and this reflection causally influences whether we harbor various implicit biases. The other is that we can actively engage in reflection on various techniques that help eradicate implicit biases, and this reflection makes a difference to whether we engage in these activities, which in turn makes a difference to whether we harbor various implicit biases. Again, it's not that agents have to actually reflect in these ways in order to be said to have indirect reflective control; but rather it's that agents are capable of such active reflection.

The control that we have over implicit bias is even more indirect than our control over our beliefs. The more indirect (or mediated) our control, the less blameworthy we are for our failures; we don't hold an individual as responsible for her implicit biases as we do her explicit ones. For example, the explicit belief that black men are more likely to commit crimes than white men, is something which we have indirect reflective control over and thus can be blamed for, given that our evidence does not support it. Voluntarily engaging in reflection on the evidence for this claim makes a difference to whether we believe it. But since reflection on the evidence for the claim doesn't necessarily make a difference to whether we harbor an implicit bias associating black men with crime, our reflective control over the bias is more indirect. It's mediated by reflecting on things not directly related to the content of the bias, such as other beliefs and values, and techniques for eradicating implicit bias. Given this mediation, it will often be the case that it's hard for agents to exercise control in these ways, and to the extent that they don't have indirect reflective control, they are not blameworthy for failing to eradicate their implicit biases.<sup>51</sup> We think this fits with our intuitions about blame for implicit bias: we are not as blameworthy for implicit bias as we are for explicit bias, because our control over implicit bias is more indirect than our control over explicit bias.

Ultimately, the question of whether we in fact have indirect reflective control over our implicit biases is an empirical one. But as we have noted, there is empirical data suggesting that with respect to many of our implicit biases, we have what amounts to indirect reflective control over whether we harbor them.

---

<sup>51</sup>There are two ways to understand the impact of an agent's lack of control over  $X$  on the agent's duty with respect to  $X$ : lack of control might entail that the agent is therefore not subject to a robust duty regarding  $X$ , or it might entail that though she is subject to the duty, she is not blameworthy for failing to fulfill it.

## 6 Conclusion

Implicit bias is widespread. We have argued that we have an epistemic duty to eradicate implicit biases with negative epistemic impact. This is because implicit bias prevents us from fulfilling other epistemic duties, and it is perhaps also intrinsically bad. But we only have this duty, it seems, if we have the right kind of control over implicit biases. That control, we argued, is indirect reflective control. If it turns out that we have this kind of control over our implicit biases—and we’ve offered empirical evidence that we do—then we have a robust epistemic duty to eradicate them, and we are thus blameworthy if we fail to do so.<sup>52</sup>

## References

- Alston, William P. 1988. “The Deontological Conception of Justification.” *Philosophical Perspectives* 2:257–299.
- Amir, Yehuda. 1969. “Contact Hypothesis in Ethnic Relations.” *Psychological Bulletin* 71:319–342.
- Azar, Beth. 2008. “IAT: Fad or Fabulous?” *Monitor on Psychology* 39:<http://www.apa.org/monitor/2008/07-08/psychometric>.
- Ballantyne, Nathan. 2015. “Debunking Biased Thinkers (Including Ourselves).” *Journal of the American Philosophical Association* 1:141–162.
- Bertrand, Marianne and Mullainathan, Sendhil. 2004. “Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination.” *American Economic Review* 94:991–1013.
- Brewer, Marilyn B and Miller, Norman. 1988. “Contact and cooperation: When do they work?” In PA Katz and DA Taylor (eds.), *Perspectives in social psychology. Eliminating racism: Profiles in controversy*, 315–326. New York: Plenum Press.
- Briñol, Pablo, Petty, Richard E., and McCaslin, Michael J. 2009. “Changing Attitudes on Implicit versus Explicit Measures: What is the Difference?” In Richard E. Petty, Russell H. Fazio, and Pablo Briñol (eds.), *Attitudes: Insights from the New Implicit Measures*. Psychology Press.
- Chisholm, Roderick. 1989. *Theory of Knowledge*. Englewood Cliffs: Prentice-Hall, 3 edition.

---

<sup>52</sup>Thanks to Andrew M Bailey, Franz-Peter Griesmaier, and Kevin McCain for comments and discussion.

- Cook, Stuart W. 1985. "Experimenting on social issues: The case of school desegregation." *American Psychologist* 40:452–460.
- Corlett, J. Angelo. 2003. *Race, Racism, and Reparations*. Cornell University Press.
- Dasgupta, Nilanjana. 2013. "Implicit Attitudes and Beliefs Adapt to Situations: A Decade of Research on the Malleability of Implicit Prejudice, Stereotypes, and the Self-Concept." *Advances in Experimental Social Psychology* 47:233–279.
- David, Marian. 2001. "Truth as the Epistemic Goal." In *Knowledge, Truth, and Duty*, 151–169. New York: Oxford University Press.
- De Houwer, Jan. 2014. "A Propositional Model of Implicit Evaluation." *Social and Personality Psychology Compass* 8:342–353.
- DePaul, Michael. 2001. "Value Monism in Epistemology." In Matthias Steup (ed.), *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*. Oxford University Press.
- Devine, Patricia G., Plant, E. Ashby, Amodio, David M., Harmon-Jones, Eddie, and Vance, Stephanie L. 2002. "The Regulation of Explicit and Implicit Race Bias: The Role of Motivations to Respond Without Prejudice." *Journal of Personality and Social Psychology* 82:835–848.
- Fazio, Russell H. 2007. "Attitudes as object-evaluation associations of varying strength." *Social Cognition* 25:603–637.
- Feldman, Richard. 1988. "Epistemic Obligations." *Philosophical Perspectives* 2:235–256.
- Firth, Roderick. 1978. "Are Epistemic Concepts Reducible to Ethical Concepts?" In Alvin Goldman and Jaegwon Kim (eds.), *Values and Morals: Essays in Honor of William Frankena, Charles Stevenson, and Richard Brandt*, 215–229. Dordrecht: Kluwer.
- Follenfant, Alice and Ric, Francois. 2010. "Behavioral rebound following stereotype suppression." *European Journal of Social Psychology* 40:774–782.
- Frankish, Keith. 2016. "Playing Double: Implicit Bias, Dual Levels, and Self-Control." In Michael Brownstein and Jennifer Saul (eds.), *Implicit Bias and Philosophy, Volume I: Metaphysics and Epistemology*, 23–46. Oxford University Press.
- Gawronski, Bertram and Bodenhausen, Galen V. 2011. "The Associative-Propositional Evaluation Model: Theory, Evidence, and Open Questions." *Advances in Experimental Social Psychology* 44:59–127.

- Gendler, Tamar Szabó. 2008a. "Alief and Belief." *The Journal of Philosophy* 105:634–663.
- . 2008b. "Alief in Action (and Reaction)." *Mind and Language* 23:552–585.
- Gregg, Aiden P., Seibt, Beate, and Banaji, Mahzarin R. 2006. "Easier done than undone: Asymmetry in the malleability of implicit preferences." *Journal of Personality and Social Psychology* 90:1–20.
- Haack, Susan. 2001. "'The Ethics of Belief' Reconsidered." In Matthias Steup (ed.), *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*. Oxford University Press.
- Hall, Richard J and Johnson, Charles R. 1998. "The Epistemic Duty to Seek More Evidence." *American Philosophical Quarterly* 35:129–139.
- Hess, Markus Patrick. 2010. *Is Truth the Primary Epistemic Goal?* Ontos.
- Hieronymi, Pamela. 2004. "The Force and Fairness of Blame." *Philosophical Perspectives* 18:115–148.
- . 2006. "Controlling Attitudes." *Pacific Philosophical Quarterly* 87:45–76.
- . 2008. "Responsibility for Believing." *Synthese* 161:357–373.
- Holroyd, Jules. 2012. "Responsibility for Implicit Bias." *Journal of Social Philosophy* 43:274–306.
- Holroyd, Jules, Scaife, Robin, and Stafford, Tom. 2017. "What is Implicit Bias?" *Philosophy Compass* 12.
- Holroyd, Jules and Sweetman, Joseph. 2016. "The Heterogeneity of Implicit Bias." In Michael Brownstein and Jennifer Saul (eds.), *Implicit Bias and Philosophy, Volume I: Metaphysics and Epistemology*, 80–103. New York, NY: Oxford University Press.
- Huebner, Bryce. 2009. "Trouble with Stereotypes for Spinozan Minds." *Philosophy of the Social Sciences* 39:63–92.
- Hughes, Sean, Barnes-Holmes, Dermot, and Houwer, Jan De. 2011. "The dominance of associative theorizing in implicit attitude research: propositional and behavioral alternatives." *The Psychological Record* 61:465–496.
- Jost, John T., Rudman, Laurie A., Blair, Irene V., Carney, Dana R., Dasgupta, Nilanjana, Glaser, Jack, and Hardin, Curtis D. 2009. "The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore." *Research in Organizational Behavior* 29:39–69.

- Kawakami, Kerry, Phills, Curtis E, Steele, Jennifer R, and Dovidio, John F. 2007. "(Close) Distance Makes the Heart Grow Fonder :Improving Implicit Racial Attitudes and Interracial Interactions Through Approach Behaviors." *Journal of Personality and Social Psychology* 92:957–971.
- Kelly, Daniel and Roedder, Erica. 2008. "Racial Cognition and the Ethics of Implicit Bias." *Philosophy Compass* 3:522–540.
- Kirwan Institute for Race and Ethnicity. 2015. "<http://kirwaninstitute.osu.edu/research/understanding-implicit-bias/>."
- Kvanvig, Jonathan L. 2005. "Truth and the Epistemic Goal." In Matthias Steup and Ernest Sosa (eds.), *Contemporary Debates in Epistemology*, 285–295. Malden, MA: Blackwell.
- Lackey, Jennifer. forthcoming. "The Duty to Object." *Philosophy and Phenomenological Research* .
- Leslie Ashburn-Nardo, Margo Monteith, Corrine Voils. 2001. "Implicit associations as the seeds of intergroup bias: How easily do they take root?" *Journal of Personality and Social Psychology* 81:789–799.
- Levy, Neil. 2015. "Neither fish nor fowl: implicit attitudes as patchy endorsements." *Noûs* 49:800–823.
- . 2016. "Implicit Bias and Moral Responsibility: Probing the Data." *Philosophy and Phenomenological Research* 93:3–26.
- Macrae, C. Neil, Bodenhausen, Galen V, Milne, Alan B., and Jetten, Jolanda. 1994. "Out of Mind but Back in Sight: Stereotypes on The Rebound." *Journal of Personality and Social Psychology* 67:808–817.
- Madva, Alex. 2016. "Why implicit attitudes are (probably) not beliefs." *Synthese* 193:2659–2684.
- . 2017. "Biased against Debiasing: On the Role of (Institutionally Sponsored) Self-Transformation in the Struggle against Prejudice." *Ergo* 4.
- Mandelbaum, Eric. 2016. "Attitude, inference, association: on the propositional structure of implicit bias." *Noûs* 50:629–658.
- McKenna, Michael. 1998. "The Limits of Evil and the Role of Moral Address: A Defense of Strawsonian Compatibilism." *Journal of Ethics* 2:123–142.
- Mitchell, Chris J., Houwer, Jan De, and Lovibond, Peter F. 2009. "The propositional nature of human associative learning." *Behavioral and Brain Sciences* 32:183–198.

- Nilanjana Dasgupta, Anthony G Greenwald. 2001. "On the Malleability of Automatic Attitudes: Combating Automatic Prejudice With Images of Admired and Disliked Individuals." *Journal of Personality and Social Psychology* 81:800–814.
- Peck, Tabitha C., Seinfeld, Sofia, Aglioti, Salvatore M., and Slater, Mel. 2013. "Putting yourself in the skin of a black avatar reduces implicit racial bias." *Consciousness and Cognition* 22:779–787.
- Peters, Uwe. 2019. "Implicit Bias, Ideological Bias, and Epistemic Risks in Philosophy." *Mind and Language* 34:393–419.
- Phills, Curtis E, Kawakami, Kerry, Tabi, Emmanuel, Nadolny, Daniel, and Inzlicht, Michael. 2011a. "Mind the Gap: Increasing Associations Between the Self and Blacks With Approach Behaviors." *Journal of Personality and Social Psychology* 100:197–210.
- Phills, Curtis E, Santelli, Alexander G, Kawakami, Kerry, Struthers, C Ward, and Higgins, E Tory. 2011b. "Reducing implicit prejudice: Matching approach/avoidance strategies to contextual valence and regulatory focus." *Journal of Experimental Social Psychology* 47:968–973.
- Rettler, Lindsay. 2018a. "Faith, Belief, and Control." *American Philosophical Quarterly* 55:95–109.
- . 2018b. "In Defense of Doxastic Blame." *Synthese* 195:2205–2226.
- Rudman, Laurie A., Ashmore, Richard D., and Gary, Melvin L. 2001. "'Unlearning' automatic biases: The malleability of implicit prejudice and stereotypes." *Journal of Personality and Social Psychology* 81:856–868.
- Ryan, Sharon. 2003. "Doxastic Compatibilism and the Ethics of Belief." *Philosophical Studies* 114:47–79.
- Rydell, Robert J. and McConnell, Allen R. 2006. "Understanding implicit and explicit attitude change: A systems of reasoning analysis." *Journal of Personality and Social Psychology* 91:995–1008.
- Saul, Jennifer. 2013a. "Implicit Bias, Stereotype Threat, and Women in Philosophy." In Fiona Jenkins and Katrina Hutchison (eds.), *Women in Philosophy: What Needs to Change?*, 39–60. Oxford University Press.
- . 2013b. "Scepticism and Implicit Bias." *Disputatio* 5:243–263.
- Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, Mass.: Harvard University Press.
- Schwitzgebel, Eric. 2010. "On Being Good at Seeming Smart."

- . 2013. “A Dispositional Approach to Attitudes: Thinking Outside of the Belief Box.” In Nikolaj Nottelmann (ed.), *New Essays on Belief*, 75–99. New York, NY: Palgrave-MacMillan.
- Shah, Nishi. 2002. “Clearing Space for Doxastic Voluntarism.” *The Monist* 85:436–445.
- Smith, Angela M. 2000. “Identification and Responsibility.” In Ton Van Den Beld (ed.), *Moral Responsibility and Ontology*, 233–246. Kluwer Academic Publishers.
- . 2005. “Responsibility for Attitudes: Activity and Passivity in Mental Life.” *Ethics* 115:236–271.
- . 2008. “Control, Responsibility, and Moral Assessment.” *Philosophical Studies* 138:367–392.
- . 2013. “Moral Blame and Moral Protest.” In D Justin Coates and Neal A Tognazzini (eds.), *Blame: Its Nature and Norms*, 27–48. Oxford University Press.
- Stapleford, Scott. 2012. “Epistemic Duties and Failure to Understand One’s Evidence.” *Principia* 16:147–177.
- Steup, Matthias. 2008. “Doxastic Freedom.” *Synthese* 161:375–392.
- . 2011. “Belief, Voluntariness, and Intentionality.” *dialectica* 65:537–559.
- . 2012. “Belief, Control, and Intentionality.” *Synthese* 188:145–163.
- Strawson, P.F. 1974. *Freedom and Resentment and Other Essays*. London: Methuen.
- Sullivan-Bissett, Ema. 2014. “Implicit Bias, Confabulation, and Epistemic Innocence.” *Consciousness and Cognition* 33:548–560.
- Wallace, R. Jay. 1994. *Responsibility and the Moral Sentiments*. Harvard University Press.
- . 2007. “Reasons, Relations, and Commands: Reflections on Darwall.” *Ethics* 118:24–36.
- . 2010. “Hypocrisy, Moral Address, and the Equal Standing of Persons.” *Philosophy and Public Affairs* 38:307–341.
- . 2011. “Dispassionate Opprobrium: On Blame and the Reactive Sentiments.” In R.J. Wallace, R. Kamar, and S. Freeman (eds.), *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon*, 348–372. Oxford University Press.
- Weatherston, Brian. 2008. “Deontology and Descartes’ Demon.” *Journal of Philosophy* 105:540–569.



Wolf, Susan. 1990. *Freedom Within Reason*. Oxford University Press.