

IMPLICIT BIAS AND THE IDEALIZED RATIONAL SELF

NORA BERENSTAIN

University of Tennessee, Knoxville

The underrepresentation of women, people of color, and especially women of color—and the corresponding overrepresentation of white men—is more pronounced in philosophy than in many of the sciences. I suggest that part of the explanation for this lies in the role played by the idealized rational self, a concept that is relatively influential in philosophy but rarely employed in the sciences. The idealized rational self models the mind as consistent, unified, rationally transcendent, and introspectively transparent. I hypothesize that acceptance of the idealized rational self leads philosophers to underestimate the influence of implicit bias on their own judgments and prevents them from enacting the reforms necessary to minimize the effects of implicit bias on institutional decision-making procedures. I consider recent experiments in social psychology that suggest that an increased sense of one's own objectivity leads to greater reliance on bias in hiring scenarios, and I hypothesize how these results might be applied to philosophers' evaluative judgments. I discuss ways that the idealized rational self is susceptible to broader critiques of ideal theory, and I consider some of the ways that the picture functions as a tool of active ignorance and color-evasive racism.

1. Introduction

In their sharp indictment of philosophy's skewed demographics, Cherry and Schwitzgebel (2016) critique the discipline for its serious diversity problem. Eighty-six percent of U.S. philosophy PhDs are "non-Hispanic white." Black philosophers account for only two percent of U.S. PhD recipients. Men represent 72 percent of PhD recipients and 80 percent of full professors. Haslanger (2013) further details how bleak the picture is with respect to philosophy's representation of women, people of color, and especially women of color. The

Contact: Nora Berenstain <nberenst@utk.edu>

percentage of tenured or tenure-track positions at ranked departments that were held by women in 2011 was just 21.9. In 2003, “the percentage of women in full-time instructional post-secondary positions was a mere 16.6 percent of the total 13,000 philosophers, a year when 27.1 percent of the doctorates went to women.” Worse yet, the 16.6 percent included no women of color, as there was “insufficient data for any racial group of women other than white women to report” (Haslanger 2013). It took a high-profile case of sexual harassment to bring these numbers to public attention. But as Haslanger (2013) notes, “With these numbers, you don’t need sexual harassment or racial harassment to prevent women and minorities from succeeding, for alienation, loneliness, implicit bias, stereotype threat, microaggression[s], and outright discrimination will do the job.”

A wide range of causal factors function to maintain the whiteness and maleness of academic philosophy. Sexual harassment and bystander silence (Jaschik 2011; Saul 2013; Schuman 2014), racist and sexist microaggressions (Sue, Capodilupo, & Holder 2008), stereotype threat (Beebee 2013; Good, Rattan, & Dweck 2012), solo status (Sekaquaptewa & Thompson 2003), and implicit bias (Jost et al. 2009) all work to destabilize and undermine scholars who are women and/or of color as well as to ensure that the overrepresentation of white men in philosophy is maintained. These factors are not unique to philosophy and negatively influence climates in other disciplines that have poor gender balance and an underrepresentation of people of color (Beattie, Cohen, & McGuire 2013; Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman 2013).

Dotson (2012) and Ruíz (2014) have also identified several methodological and cultural norms within philosophy that function to reify certain modes of thought as the domain of philosophy proper while excluding from the discipline alternative ways of engaging philosophically with the world. Dotson critiques what she calls academic philosophy’s “culture of justification,” namely, its preoccupation with legitimation narratives that requires the constant defense of those projects and practices that, while philosophical in nature, are not widely recognized as such within the discipline (2012: 5). The norms and practices that this culture comprises function not only to promote the general whiteness and maleness of the discipline but to create an intellectually hostile environment for women of color philosophers and for diverse practitioners of philosophy more generally. Ruíz emphasizes that this culture of justification quickly becomes a culture of alienation for those that are disproportionately subject to its demands. She writes, “Not only do the institutionalized norms and disciplinary practices of academic philosophy regularly collide with the intersectionally complex being-in-the-world of women of color, they do so in a way that produces experiences of normalized alienation” (2014: 197). Philosophy’s culture of justification and the systematic microaggressions that such a culture licenses against marginalized

and diverse practitioners exemplify how methodological norms and practices in the discipline create an environment of exclusion.

Are there other methodological factors specific to philosophy that are contributing to the more significant overrepresentation of white men in the discipline?¹ I identify a specific set of disciplinary methods that function alongside the norms that Dotson and Ruíz characterize to uphold practices of exclusion within academic philosophy. Specifically, I suggest that certain of philosophy's theoretical and methodological commitments create a unique set of obstacles when it comes to addressing implicit bias in disciplinary decision-making practices. An idealized notion of the rational self and a corresponding reliance on introspection as a reliable method of knowing one's own attitudes are two factors that are relatively more prevalent in philosophy than in other disciplines. I suggest that these methodological and theoretical features of philosophy allow inflated levels of skepticism about the existence of bias to flourish among members of the discipline and that this skepticism functions to prevent departments from taking necessary steps to minimize the influence of implicit bias on institutional decision-making procedures.

A few salient facts provide informal evidence that philosophy demonstrates greater resistance than other academic disciplines to instituting practices and reforms necessary to reduce effects of implicit bias on institutional decision-making. One is that the American Philosophical Association (APA) did not release its "Good Practices Guide" until 2017 (APA Task Force on A Best Practices Guide 2017). Among many important topics related to reducing the discipline's hostility to marginalized groups, the Good Practices Guide includes a section on countering implicit bias. Before 2017, the APA did not officially endorse any specific measures for reducing the effects of bias in departmental hiring and admissions decisions.

Contrast this, for instance, with the fact that the American Psychological Association released a statement on recruiting and hiring ethnic minority faculty that included a significant section on bias as early as 1996 (APA Commission on Ethnic Minority Recruitment, Retention, and Training in Psychology 1996). The American Psychological Association made efforts to address these issues more than 20 years before the American Philosophical Association did. The discipline of psychology seems to have generally made more tangible and material

1. The only STEM fields with lower numbers of women than philosophy are computer science, engineering, and physics (Saul 2011). The overwhelming whiteness of philosophy is evident in the racial demographics of those who received doctoral degrees in the field in 2009: 92% white, 2.5% Black, 0.5% Native American, 3% Asian or Pacific Islander, and 2% Hispanic, approximately (APA 2013). Botts, Bright, Cherry, Mallarangeng, and Spencer (2016) put the percent of affiliated U.S. philosophers who are Black at 1.32%. This number includes those with a PhD who are employed by a philosophy department as well as current PhD students.

institutional efforts to enhance racial and gender diversity in the discipline than philosophy has. As far back as 1998, the American Psychological Association Board of Directors and Council of Representatives dedicated \$27,000 to promote recruitment and retention of psychologists of color and ethnic minorities. The American Psychological Association increased this amount every year until it reached \$100,000 in 2004 (Suinn et al. 2005).

Even the discipline of mathematics seems to be ahead of philosophy in addressing issues of bias. The Mathematical Association of America (MAA) released its official guidelines for avoiding implicit bias in selection of recipients for MAA awards for scholarship, writing, teaching, and service in 2012, five years before the American Philosophical Association endorsed any such official guidelines for minimizing bias in decision-making (AWIS-MAA Joint Task Force on Prizes and Awards 2012). While the MAA prepared these guidelines for guiding selection of prize recipients, they explicitly acknowledge their relevance to committee decisions of all sorts.

Hermanson (2017) cites Crouch and Schwartzmann (2012) as crediting Haslanger (2008) with bringing the issue of implicit bias to the attention of the philosophical community. Hermanson's article also offers an example of a broader trend within the discipline, as he himself is skeptical that implicit bias plays any explanatory role whatsoever in the underrepresentation of women in the philosophy. He questions whether Moss-Racusin et al.'s (2013) study showing a high level of gender bias in judgments of CV quality in psychology can be taken as even *prima facie* evidence that implicit bias plays a role in philosophy hiring. He also expresses the view that the "zeal with which Departments are promoting diversity initiatives" (presumably including steps to reduce the effects of bias) is merely a matter of "political correctness" (Hermanson 2017: 6). Hermanson's claims are not unique.

The tenor of many public debates in online philosophical forums suggests a non-negligible level of resistance within the discipline (and at the very least, a high level of resistance among the discipline's most vocal members) to enacting measures to minimize the effects of bias. This is evident, for instance, in comments on Leiter (2014) and Leiter (2016). On a Leiter (2016) post about Jennings, Cobb, and Vinson's (2016) data showing that women were more likely than men to secure a permanent position in philosophy within two years of getting their PhD, one anonymous commentator suggests that bias in favor of hiring women is a likely explanation for why. They write, "Somehow that is not on the table as a serious option, although it is obvious to anyone in the profession that this is a front-runner as a hypothesis." Another commentator emphasizes how challenging it is to be "white and male" on the philosophy job market. On another Leiter (2014) post about a controversial article suggesting there is no reason to believe in discrimination against women in philosophy (Sesardic & de Clercq 2014), a

pseudonymous commentator writes, “Affirmative Action certainly seems, if anything, to have overcorrected for any implicit bias, giving women real advantages over men, and on the matter of greatest importance in a career: securing positions in a highly competitive environment.” Later in the thread, the same commentator acknowledges being “less familiar with the literature on implicit bias.” These comments are suggestive of a view that many seem to hold in philosophy, namely that there is little if any implicit bias against members of marginalized groups and that explicit bias in fact favors members of marginalized groups, giving them a significant advantage over members of dominant groups, such as white men, in hiring and other domains.

While implicit bias is pervasive within academia, I conjecture that certain factors make implicit bias a particularly powerful phenomenon in philosophy and thus help to explain its relative lack of gender and racial diversity.² I propose that the concept of the idealized rational self acts as an obstacle to achieving an accurate and complete understanding of implicit bias. I focus on two possible effects of embracing the idealized rational self in philosophical theorizing:

- 1). It makes thinkers skeptical of the correct picture of implicit bias in general.³
- 2). It causes thinkers to be skeptical that they themselves are subject to implicit bias and to underestimate the extent to which their judgments are affected by it.

I suggest that philosophers who employ a conception of the idealized rational self are less likely to give serious consideration to the existence and impact of implicit bias in general and on their own reasoning and are thus less likely to

2. It must be emphasized that the role of implicit bias in explaining philosophy’s whiteness and maleness is partial at best. Implicit bias is only one of the many ways that larger social structures of domination operate within the discipline. Discussions of implicit biases make little sense when separated from the context of interlocking structures of oppression that produce such biases. Some discussions of implicit bias treat it as the sole determining factor for philosophy’s lack of diversity. This ignores and obscures the myriad ways that academic philosophy perpetuates and maintains white supremacy, sexism, cisheteropatriarchy, ableism, classism, and other structures of oppression. For an explanation of how attributing the underrepresentation of Black philosophers in the discipline solely to implicit bias obscures professional philosophy’s “overwhelming, functionally exclusionary whiteness,” see McPherson (2011). This paper aims to recognize the role that implicit bias plays in perpetuating philosophy’s exclusion of certain underrepresented groups while situating it within the larger contexts of structural oppression in which bias operates.

3. That is, it causes such thinkers to doubt the existence or pervasiveness of introspectively non-transparent biases generally affecting members of a culture in which there are pervasive prejudices against certain groups, and specifically affecting members of a discipline that is stereotyped a certain way—e.g, as being paradigmatically represented by someone who is both white and male. See Beebee (2013) for further explication of occupational stereotypes regarding who a ‘typical’ philosopher is.

take the necessary steps to combat the effects of implicit bias on their reasoning and decision-making.⁴ This hypothesis has broadly empirical dimensions and is motivated in part by empirical work in the social sciences. A primary goal of this paper is to show that this hypothesis warrants further experimental investigation and that such investigation would pragmatically benefit the discipline of philosophy.

In the following section, I explicate the view of the idealized rational self and provide examples of its descendants in contemporary philosophy. Section 3 develops the hypothesis after reviewing some of the empirical literature on implicit bias and its effects in higher education. Section 4 surveys suggested best practices for hiring and publishing based on empirical data on implicit bias. In Section 5, I consider recent experiments in social psychology that suggest that an increased sense of one's own objectivity leads to greater reliance on bias in hiring scenarios. I hypothesize ways these results might be applied to philosophers who conceive of themselves as generally rational and objective. In Section 6, I apply Mills's (2005) criticisms of ideal theory to the idealized rational self and I consider some of the ways that the picture interacts with the phenomena of active ignorance and color-evasive racism.

2. The Idealized Rational Self and its Philosophical Vestiges

While many philosophers take the discipline to be broadly continuous with the sciences, at least in principle, much of philosophy remains insulated from contemporary developments in the sciences. Some concepts that are considered antiquated and obsolete in cognitive science remain operative within philosophy, even functioning as cornerstones of philosophical theorizing. Without directly addressing their viability, I consider a specific conception of the self and a corresponding set of methodological assumptions that have relatively more influence in philosophy than they do in the sciences.⁵

4. At the departmental and institutional levels, the unwillingness to take action to undermine the effects of implicit bias on decision procedures may translate into distorted evaluations at every stage of the pipeline (Paxton, Figdor, & Tiberius 2012).

5. Since this aims to explain why philosophy's levels of racial and gender diversity are worse than those in many of the STEM disciplines, what matters primarily is that such views gain more traction in philosophy than in scientific disciplines with better representation of marginalized groups. These include molecular biology, biochemistry, chemistry, and mathematics (Healy 2011). (Note that these numbers refer only to the percentage of women in various academic disciplines. They do not address the percentages of women of color in these disciplines nor their percentages of people of color more generally.) While it is true that some of these disciplines probably have no general notion or understanding of the nature of mind and rationality, having no such conception may well be less of a barrier to understanding and addressing implicit bias than having a concep-

The main concept that I focus on is what I refer to as the idealized rational self. This idealization takes shape differently across various areas of philosophy. It can be found in numerous sub-disciplines from epistemology to moral philosophy. At its core, the commitments of the idealized rational self are that “rationality alone is essential to the self and that the ideal self is transparent, unified, coherent, and independent” (Meyers 2010). I suggest that the self, on this picture, is marked by the following features:

- i. Consistency – the mind does not contain contradictory beliefs;
- ii. Unity – the mind is simple; it does not have parts and is not subdivided;
- iii. Rational transcendence – in principle, thinkers are capable of achieving an objective rationality that abstracts away from their contingent circumstances and transcends cultural, environmental, and bodily limitations on reasoning (e.g., exhaustion, hunger, and pain);⁶
- iv. Introspective transparency – the contents of one’s mental states are transparently available through introspection.

Aspects of the idealized rational self can be found in a number of philosophical domains, both contemporary and historical. For one illustration of the conception of the idealized rational self, consider the following passage from Descartes:

There is a great difference between the mind and the body. Whereas every body is by its nature divisible, the mind can’t be divided. For when I consider the mind, or consider myself insofar as I am merely a thinking thing, I can’t detect any parts within myself; I understand myself to be something single and complete. The whole mind seems to be united to the whole body, but not by a uniting of parts to parts, as the following consideration shows.

If a foot or arm or any other part of the body is cut off, nothing is thereby taken away from the mind. As for the faculties of willing, of understanding, of sensory perception and so on, these are not parts of the mind, since it is one and the same mind that wills, understands and perceives. They are (I repeat) not parts of the mind, because they are properties or powers of it. By contrast, any corporeal thing can easily be divided into parts in my thought; and this shows me that it is really divisible. This one argument would be enough to show me that the mind is completely dif-

tion of rationality that represents deliberative processes as primarily a priori, sensitive to reasons, and readily introspectable.

6. This phrasing is drawn from language in (Meyers 2010).

ferent from the body, even if I did not already know as much from other considerations. (1996: 58, Sixth Meditation)

Descartes differentiates the mind from the body by denying that the mind has parts, which suggests an acceptance of unity. By claiming that nothing is lost from the mind when part of the body is removed, Descartes shows his adherence to rational transcendence. His argument relies on introspective transparency, as the unity of the mind seems to follow from the fact that one cannot detect any parts of one's mind simply by considering it. Further, we know that Descartes accepts introspective transparency because of his belief that an evil demon could not deceive him about the contents of his own mind. On this picture, the mind is not subject to the limits of the body and we can know this by simply turning the mind on itself. Descartes' philosophy of mind exemplifies the most robust version of the idealized rational self conception.

Though philosophers now widely reject the Cartesian theory of mind, vestiges of this picture remain influential in contemporary debates in ethics, epistemology, moral psychology, and action theory. Because conceptual offshoots of the view continue to appear within philosophical debates, even among those who would explicitly reject the robust Cartesian picture of the idealized rational self described above, the focus of this paper is not solely on the hypothesized effects of embracing the picture of the idealized rational self in its entirety. My project bears some similarities to Gehrman's (2016) compelling argument that ethics continues to be both influenced and limited by an outdated hierarchical conception of the natural world, often referred to as The Great Chain of Being, which places humans at the top of the chain closest to God. Gehrman suggests that while the Great Chain conception of nature has been roundly rejected in light of evolutionary theory, and that many moral philosophers eschew any reference to the divine, vestiges of the view continue to influence the philosophical domains in which it was once prized. Gehrman sees her own project as parallel in structure to Anscombe's (1958) argument in "Modern Moral Philosophy" that certain widely used moral concepts are inherited from divine law ethics and make little sense on their own once their origin has been removed from moral discourse. She writes,

Like divine law ethics, the Great Chain of Being is a conception that was dominant for many centuries. And even though in many ways it has been given up, it is, as Anscombe says, only 'a natural result' of its long-time dominance that certain elements of it should remain to organize and constrain human thought in the relevant domains, 'though they have lost their root'. (Gehrman 2017: 1, quoting Anscombe 1958: 2)

My project shares Gehrman's and Anscombe's goal of identifying "survivals, or derivatives from survivals, from an earlier conception of ethics which no longer generally survives." (Gehrman 2017: 1, quoting Anscombe 1958: 1). However, in the case of the idealized rational self, the domains of the earlier conceptions and remaining vestiges are not limited exclusively to ethics. This paper thus takes into account explicit avowals of the idealized rational self as well as philosophical conceptions of rationality, consciousness, and selfhood that contain elements descended from the conception's earlier iterations.

For one example of a philosophical debate that retains elements of the idealized rational self picture, consider how an adherence to rational transcendence is implicit in many contemporary views on the methodological use of intuitions in analytic philosophy. While the use of intuitions in philosophy originally began as an analogue to the practice within Chomskian linguistics of using a native speaker's intuitions as evidence of a sentence's grammaticality, it has since expanded to treating intuitions as defeasible evidence about everything from the nature of truth to the relationship between rightness and goodness (Symons 2008). Bealer (2000: 4) describes intuitions as intellectual seemings that are "inherently more resistant" than belief to contingent influences. In his explanation of the relationship between intuition and concept possession, Bealer employs several illustrative examples which he emphasizes are "designed so that neither features of other people nor of the larger social or linguistic context are relevant. Nor are features of the environment" (2000: 12). Intuitions, while defeasible and revisable, are viewed as unmediated by contingencies such as the cultural or social environment of the intuiter.⁷ Bengson (2015), Markie (2013), Huemer (2007), and Korman (2005), among others, argue for the justificatory power and evidential value of intuitions. Sosa (2007) asserts that intuitions do more than merely tell us about our concepts, they tell us about the nature of things themselves. This methodology suggests that thinkers are capable of achieving objectivity through the use of intuition, a rational faculty that is unaffected by contingent circumstances related to embodiment or environment. A commitment to rational transcendence is thus evident in the areas of analytic philosophy that place significant epistemic trust in the presumed rational faculty of intuition.

Another example of an offshoot of the idealized rational self in a prominent analytic debate is Kripke's (1979) puzzle about belief, in which consistency plays a central role. The presumption that rational agents have only consistent

7. Intuitions are also, for Bealer, by and large consistent, which further suggests that his view is influenced by a picture of the idealized rational self. Bealer writes,

To be sure, the logical paradoxes and other antinomies have shown that certain intuitions can be inconsistent. But this pales by comparison with a positive fact, namely, the on-balance consistency of our elementary concrete-case intuitions. Indeed, the on-balance consistency of our elementary concrete-case intuitions is one of the most impressive general facts about human cognition. (2000: 4)

beliefs is required for the set-up of Kripke's original puzzle and is retained in much of the contemporary literature that has followed. The widely discussed puzzle arises from direct-reference theory, also called Millianism, which supposes that a proper name contributes no semantic content beyond its referent to a sentence that contains it. There are numerous versions of Kripke's puzzle. I consider the stripped-down version in which substitution of co-referring names plays no role. The version is as follows: Peter, looking at a picture of the famous pianist Paderewski assents to the claim that 'Paderewski had musical talent.' Later, Peter reads a newspaper article about Paderewski, the Polish statesman. Unbeknownst to Peter, Paderewski the pianist and Paderewski the statesman are one and the same. Since Peter doubts the musical abilities of politicians, he assents to the claim 'Paderewski had no musical talent.' If we assume the truth of the disquotational principle—that if a speaker assents to 'p' then the speaker believes that p—then it appears that Peter holds contradictory beliefs. In response to the conclusion that Peter both believes that Paderewski had musical talent and believes that Paderewski had no musical talent, Bach writes, "But this seems incorrect. After all, Peter is not being illogical – he does not believe contradictory things – he is merely ignorant" (1997: 224). Bach presumes at the outset that Peter should not be taken to have contradictory beliefs, since he is "merely ignorant" rather than "illogical" or really, irrational. Sosa's (1996: 380) reconstruction of the arguments in Kripke (1979) makes this assumption explicit in a premise: "(9) If Peter has contradictory beliefs, then Peter is not rational." In order to generate the puzzle at all, it is necessary to assume that Peter is rational and that rationality requires not believing contradictory things.

Bach discusses a number of different solutions to the puzzle, none of which involve relinquishing the assumption that rational people do not hold contradictory beliefs. Some philosophers have argued that the puzzle cases are ones in which the agent in question "rationally believes each of a pair of straightforwardly contradictory, or at least impossible, propositions" (McGlone 2009: 502). McGlone (2009) notes that Braun (1998; 2006), Crimmins (1992), and Richard (1990) among others all argue for or are at least sympathetic to this view. But many others argue against it. These include Taschek (1998), Bach (1997), Sosa (1996), Forbes (1990), Lewis (1981), Marcus (1981), and Stalnaker (1981; 1987). Some philosophers, such as Sosa (1996), explicitly state that holding contradictory beliefs entails irrationality. Sosa even suggests that such intuitions about rationality and epistemic normativity were among Kripke's central motivations in imagining the cases. In describing the puzzles, Sosa writes, "we have a presumably rational agent who, according to certain seemingly innocent principles, is characterized as inconsistent and subject to serious epistemic censure" (1996:

378).⁸ For Sosa, failing the criterion of consistency opens one to severe disapprobation. The presumption that rational minds are consistent and do not contain contradictory beliefs seems to be doing much of the work not only in generating the original puzzles but in maintaining philosophers' continuing puzzlement. Notably, direct-reference theory is usually considered to be the more deflationary of the two primary competing views on names, while Fregeanism is the view more closely associated with a Cartesian picture of mind. Yet as Kripke's puzzle shows, the idealized rational self influences debate even among those who accept the deflationary Millian view of proper names and reject the Fregean view as spooky or mysterious because of its preservation of ineffability in mental and linguistic content. The literature on Kripke's puzzle about belief demonstrates that an assumption that the mind is consistent is alive and well even beyond Cartesian circles.

The final example of a descendent of the idealized rational self view that I discuss is the conception of the 'single unified agent' in the philosophy of action. Consider, for instance, the picture of rationality that Scanlon assumes as the foundation for his theory of moral obligation:

When a rational creature judges that the reasons she is aware of count decisively against a certain attitude, she generally does not have that attitude, or ceases to have it if she did so before—ceases to feel conviction in regard to the belief or to use it as a premise, or ceases to look for ways to implement the intention, and is not inclined to act on it. (Scanlon 1998: 24)

According to Scanlon, the unified rational agent does not hold beliefs or attitudes that she recognizes to be unsupported by reasons and does not act or make decisions on the basis of these attitudes. Rationality, for Scanlon, implies a unity of attitudes and a responsiveness of these attitudes to reasons. He continues:

Although rational creatures commonly form beliefs, intentions, and other attitudes unreflectively, the formation of these attitudes is generally constrained by the general standing judgments about the adequacy of reasons. For example, if a person holds that a certain class of putative evidence is not good grounds for forming beliefs, or that certain reasons are not good grounds for action of a given kind, then she generally does not unreflectively form beliefs on the basis of such evidence or unreflec-

8. An assumption of unity also appears to be playing a role here, since it would otherwise be possible for each of the two beliefs that form the contradictory pair to be contained in separate 'parts' of the mind, thus avoiding inconsistency. A presumption of the unified whole thus helps to produce the apparent inconsistency in the cases.

tively take action of the given kind on the basis of those reasons. (Scanlon 1998: 24)

The rational agent does not form beliefs, intentions, or other attitudes on the basis of what she recognizes to be bad evidence. Not only do rational agents supposedly not have these beliefs, they also do not act on them. This suggests a presumption toward the unity and consistency of the attitudes of well-functioning rational agents. This sort of picture seems to be in tension with what we know about implicit bias. Contrary to what Scanlon suggests, people generally do have implicit biases that they do not consciously endorse (Staats 2014). The Kirwan Institute defines implicit biases as “attitudes or stereotypes that affect our understanding, actions, and decisions in an unconscious manner” (Staats 2014: 16). These biases encompass both positive and negative evaluations and are often activated without a subject’s awareness or control. Further, when agents are subject to implicit biases, they frequently form attitudes such as beliefs and intentions on the basis of stereotypes that they recognize to be false. In other words, agents sometimes form implicit associations on the basis of stereotypes whose content they do not and would not explicitly endorse upon reflection. An agent may hold “that a certain class of putative evidence is not good grounds for forming belief” and still unreflectively form beliefs on the basis of it. Such attitudes are capable of affecting agents’ decisions and behaviors (Amodio & Divine 2006; Green et al. 2007; McConnell & Liebold 2001; Staats 2014). Scanlon’s claim that the rational agent generally does not hold an attitude when she judges that the reasons she is aware of count against having it is thus complicated by this feature of implicit attitudes.

Scanlon’s characterization of irrationality describes exactly the situation we find ourselves in when our implicit biases conflict with our expressed beliefs and moral commitments. He writes, “Irrationality in the clearest sense occurs when a person’s attitudes fail to conform to his or her own judgments” (Scanlon 1998: 25). What we know about implicit biases is that their contents often directly conflict with agents’ conscious judgments and explicitly held values. Implicit biases certainly seem like good candidates for irrational attitudes. So what should the rational agent should do about such attitudes, given that they tend to dwell below our conscious awareness? For Scanlon, the moral obligations we have to one another exist within the space of reasons. Implicit biases tend to operate outside the space of reasons, since they are not the sort of attitudes that are generally responsive to reasons. This makes it challenging for the view to conceive of what our moral obligations are with respect to addressing issues of implicit bias, as there is no clear path forward for addressing failures of moral obligation that occur within the realm of the unconscious or the sub-doxastic.

While Scanlon's view can be understood as a normative ideal that governs rationality rather than as a descriptive generalization about deliberative practices, this does not preclude the view from obscuring important non-ideal aspects of human cognition. Section 6 further addresses why treating some elements of ideal theory as mere idealization does not necessarily prevent them from concealing significant features of the actual world in a way that produces harmful effects. Many of the concerns raised there are applicable here as well. This discussion of Scanlon is not intended to single his view out as an exceptionally strange or fringe account of rational agency. The view of rationality discussed above that Sosa (1996) endorses and attributes to Kripke (1979) might be considered extreme, as a view on which holding a pair of contradictory beliefs without realizing it entails irrationality would likely make most agents turn out to be irrational. In contrast, I take Scanlon's view to be both widely held in philosophy and significantly more workable than these more restrictive views. While Scanlon does explicitly recognize that common forms of moral bias exist and that rational agents are sometimes irrational, the way he frames rational agency incentivizes agents to downplay the extent to which they act outside of the bounds of rationality by holding attitudes or performing actions that are not appropriately responsive to reasons. Scanlon frames rational agency in terms of what the rational agent generally does or does not do, locating rationality within agents themselves. As with Sosa and Kripke's views, the fact that Scanlon frames these constraints in terms of features of rational agents as opposed to features of rational actions or rational beliefs creates tension for thinkers who wish to take themselves to be rational while also acknowledging the limits and constraints on their rationality. Framing the discussion in terms of rational actions and rational beliefs rather than rational agents could leave more room for individuals to consider themselves to be rational agents while still recognizing that they hold attitudes and perform actions that are unsupported by reasons and that result from pervasively held implicit biases.

3. The Idealized Rational Self as a Roadblock to Understanding Implicit Bias

Scanlon's conception of the unified rational agent illustrates how the notion of the idealized rational self, when paired with the common but inaccurate picture of bias outlined below, may prevent agents from acknowledging and addressing their own biases. Saul sketches the misleading picture of bias and explains what sort of conceptual change is required to begin to recognize and overcome the effects of implicit bias on our reasoning:

The first step is to move away from a widely-held picture of biases. On this conception, things like sexism and racism are a matter of conscious beliefs, easily introspectable. This is a reassuring picture, as most of us well-meaning people can introspect and see that we don't have any of those nasty racist and sexist beliefs. Having done due diligence in this way, we can trust ourselves to assess work, students and job candidates fairly. What we know about implicit biases shows this to be wrong. We are all likely to hold biases of which we are unaware, and we are likely to be similarly unaware of the ways that these biases are affecting our judgments in particular cases. Once we acknowledge this, we can begin to do something about it. (2015)

Saul points out that many agents' false negatives about the existence of their own biases are due to an overestimation of the range of attitudes that are "easily introspectable" or introspectively transparent.⁹ Introspective transparency is the feature that some attitudes have that makes them directly and immediately first-personally available, upon reflection, to the agent that holds them. Introspective transparency is often conceived of as what links the content of an agent's attitudes with their immediate conscious awareness, positioning their attitudes to be acted upon. A tenet of most analytic philosophy of mind is that agents have privileged, first-personal access to (at least some of) their own mental states and that this privileged access is produced by introspective transparency.¹⁰

The strict sense of introspective transparency invoked here must be distinguished from the broader notion of introspective access. While introspective transparency involves direct non-inferential access to the content of our mental states, introspective access applies to any mental states of which we can have first-personal knowledge, whether direct or inferential. This latter notion of introspective access can thus be applied even to views such as Carruthers's (2011) that are skeptical of the faculty of introspection. Consideration of Carruthers's skeptical view will help clarify the ways in which claiming the content of a men-

9. A presupposition of introspective transparency likely also plays a role in Scanlon's assumption that the rational agent does not have attitudes supported solely by what they recognize to be bad evidence.

10. For further discussions of introspective transparency, see Cullison (2007) and Fernández (2003). Fernández (2003) characterizes privileged access in terms of a non-empirical basis for self-ascriptions of belief, and a distinction between kinds of justification for beliefs about one's own beliefs and beliefs about others' beliefs. He writes, "When a mental state such as a sensation, perceptual experience, or a belief is self-ascribed by S, the self-ascription normally enjoys a characteristic entitlement or epistemic right" (Fernández 2003: 353). Engelbert and Carruthers (2010) document the widespread assumption in philosophy that introspection is a reliable source of self-knowledge and contrast it with the increasingly popular view among cognitive scientists that introspection is either nonexistent or unreliable. See Carruthers (2011) for arguments against the viability of introspection.

tal state is introspectively transparent is stronger than claiming that it is merely introspectively accessible.

Carruthers (2011) presents a view of mindreading on which we have no special non-interpretive access to our own attitudes. Rather, we gain knowledge of our own attitudes in much the same way that we gain knowledge of the attitudes of others—through interpretation. For Carruthers, the primary distinguishing feature in the case of self-knowledge is that we have interpretive access to a much greater amount of information on the basis of which to attribute such knowledge. We do not, however, have any special or distinctive non-interpretive access to our own attitudes. On this view, our own attitudes are not introspectively transparent, though we may still achieve knowledge of them through other means.

While our implicit attitudes are not introspectively transparent, we can still access them in the same way that Carruthers thinks we access all attitudes, via interpretation of sensory information. We may be able to accurately predict our results on implicit association tests, but this is not because we have any special direct or non-interpretive access to our implicit attitudes. If we accurately predict our own results on an IAT, for instance, it is through the same processes that allow us to accurately predict others' attitudes—interpretation of sensory input, bodily responses, and behavior. Because we can sometimes make inferences about the contents of our implicit biases from our bodily responses to various stimuli, it is fair to say that the contents of our own biases are not completely sealed off from our awareness. This claim seems to be supported by recent empirical work on implicit bias (Madva 2017).

Madva (2017) cites several studies whose results he suggests are difficult to interpret without the assumption that people are at least somewhat aware of their implicit biases. One is Cooley, Payne, Loersch, and Lei's (2015) study, which shows that telling participants that their gut feelings either do or do not represent their genuine views affects their self-reported attitudes of explicit prejudice. Participants completed the Affect Misattribution Procedure in order to measure their implicit responses to photos of gay male couples and straight male-female couples. Afterward, all participants were told that they may have had a "gut feeling" toward the pictures. They were told either that "research has shown that this gut feeling usually reflects [or does NOT reflect] people's genuine attitude toward homosexuality" (Cooley et al.: 106) and were asked to write down 2-3 reasons why the feelings they felt during the task are [or are NOT] their own attitudes. Participants then completed the Modern Homophobia Scale (MHS) to assess their explicit attitudes (Raja & Stokes 1998). There were generally high levels of implicit homophobic bias among the participants. The researchers found that those who were told their gut feelings did represent their genuine attitudes expressed significantly higher levels of explicit homophobic

prejudice in their MHS assessment than did those who were told that their gut feelings were not representative of their genuine attitudes. This suggests that participants did have some level of awareness of what their gut responses were.

Madva notes that this study and others “are difficult to interpret without supposing that participants tend to have some introspective awareness of their implicit biases” (2017: 7). While I agree with this claim, note that the evidence of ‘introspective awareness’ of biases that these studies provide is not evidence that agents have direct, non-inferential awareness of their implicit biases. A Cooley-study participant with homophobic bias might feel a negative bodily sensation when looking at a photo of a gay male couple and become aware of their bias through the interpretation of that sensation. The results of the study are perfectly in line with a Carruthersian view on the kind of access we have to our implicit attitudes. That people are sometimes able to accurately predict and report their implicit attitudes is fully compatible with the view that we do not have direct non-interpretive awareness of our implicit attitudes. A failure of introspective transparency in the case of implicit biases does not entail that implicit biases are, to use Madva’s phrase, “permanently concealed from our conscious minds” (2017: 6). They are available to us in the way that Carruthers suggests all attitudes are—through interpretation.

While implicit biases are not completely sealed off from conscious awareness, they are not transparently available to agents the way explicit biases are (Kelly & Roedder 2008). This is important to recognize, since being too permissive in the ascription of introspective transparency can result in high-cost errors. Including implicit biases in the domain of the introspectively transparent, for instance, is likely to produce unfounded skepticism about the extent to which one’s own implicit biases affect one’s decision-making. Consider the view that processes of deliberation are introspectively transparent, so that the reasons for which one acts or makes certain evaluative judgments are always introspectively available to the agent. Imagine that someone who accepts unity, consistency, and introspective transparency about deliberative processes reflects on their own decision-making procedures in an evaluative scenario. Since implicit biases are not introspectively available, they find upon reflection no evidence that bias played a role in producing their judgment. Instead they find that they consciously made an effort to be fair and base their judgment on reasons. Their belief in the idealized rational self persuades them that they would have direct first-personal knowledge of their implicit biases and the role that these biases played in producing their judgment if the bias did indeed exist and played any such role. They do not have any such knowledge. Thus, they conclude that either the biases were never there to begin with or that they were able to overcome them to produce an accurate evaluative judgment. This latter conclusion is further facilitated by a commitment to rational transcendence, and the effects of the

misattribution of introspective transparency are exacerbated by the commitment to unity.

A major problem with the picture of the unified self is that it seems not to allow for agents to hold sub-conscious attitudes that directly contradict their explicit beliefs. Belief in unity of mind combined with belief in consistency rules out the possibility that we could be genuinely affected or even primarily motivated by sub-doxastic states that conflict with our consciously held attitudes. The view of the mind as unified and consistent combined with a belief that introspection is a reliable and appropriate method for discovering one's own beliefs and mental states precludes a robust understanding of the phenomenon of implicit bias and prevents awareness of the extent to which one is affected by implicit bias. Since implicit biases are not introspectively transparent to us, those who hold that the mind is unified and consistent and that mental states are directly available to us upon reflection would mistakenly conclude from the fact that they "looked inside themselves" and found none of those aforementioned nasty racist or sexist beliefs that they were free of racial and gender-based biases. In this way, a commitment to unity combined with an overestimation about the range of attitudes that are introspectively transparent leads to an underestimation of one's own personal biases.

A commitment to rational transcendence can produce a similarly harmful optimism about the extent to which agents can resist their own susceptibility to implicit bias. Recall that rational transcendence is an agent's ability to overcome cultural, environmental, and bodily limitations on reasoning. The characterization of the self as an agent that is capable of rational transcendence stands in opposition to a view on which reasoning and judging frequently invoke heuristics and biases largely because they are more efficient than "costly" but more precise forms of reasoning (Fiske 1998; Reskin 2000). Certain situational factors greatly increase the influence of implicit bias. Subjects rely more on stereotypic preconceptions when their ability to engage in high-effort cognitive processing is reduced, as when judgments are rendered during non-optimal times of day¹¹ (Bodenhausen 1990), or when distractions are present (Petty, Wells, & Brock 1976). Individuals rely more on racial bias in hiring decisions with ambiguous criteria that require greater discretion in judging candidate qualifications (Dovidio & Gaertner 2000). Accepting the reality that situational factors affect reasoning and attending to the specific physical and environmental aspects of deliberative aspects are necessary for responding appropriately with changes to procedure.

One reason to be concerned about the influence of the idealized rational self in philosophy is that the view likely magnifies philosophers' already high lev-

11. As characterized by circadian variations in arousal level.

els of confidence in the quality of their own reasoning. In Section 5, I explore some of the ways that philosophers might be particularly susceptible to falling prey to unfounded epistemic arrogance about their vulnerability to implicit bias, and I outline some of the potentially harmful effects of philosophers' high levels of trust in their own rational judgments. For instance, I consider the possibility that the better-than-average effect is more pronounced among philosophers and that their widespread overestimation of their ability to overcome biases prevents necessary reforms from taking place. I also consider the danger that the a priori methods of philosophy and philosophy's general self-conception as a rational and objective discipline serve to consistently reinforce philosophers' sense of their own objectivity. This is significant in light of research linking a high sense of one's own objectivity to a reliance on gender bias in mock hiring scenarios. Overall, I think there is reason to be concerned that influence from the idealized rational self picture exacerbates the potentially detrimental effects of a disciplinary-wide overconfidence in philosophers' ability to resist the influence of biases on their reasoning.

4. Implicit Bias and Institutional Decision-Making

Errors that lead to false negatives about the existence and magnitude of one's own biases come with high costs. Beliefs or attitudes that make us less likely to accept that we ourselves are liable to make inaccurate judgments due to implicit bias act as barriers to changing the decision-making procedures that allow implicit bias to flourish. Given that implicit bias has the potential to negatively affect the advancement of philosophers who are not white and male at every stage of their education and career, the failure to address even marginal levels of implicit bias could lead to disproportionate effects down the line. Long before any hiring decisions are made, implicit bias may function to distort evaluative judgments when professors write letters of recommendation for their undergraduates who are women and/or of color (LaCroix 1985), when admissions committees review these students' files, when prospective students contact professors about possible future mentoring (Milkman, Akinola, & Chugh 2015), when professors place write-ups of their initial impressions in graduate students' files (Ross et al. 2017), when committees award essay prizes and fellowships to graduate students, when students are offered encouragement to submit articles to conferences and journals, and when professors write and/or read letters of recommendation for graduate students' post-doc and job applications (Dutt, Pfaff, Bernstein, Dillard, & Block 2016; Madera, Hebl, & Martin 2009; Morgan, Elder, & King 2013; Schmader, Whitehead, & Wysocki 2007; Trix & Psenka 2003). If a department's consensus is that its faculty are not particularly susceptible to implicit bias, their

default position is likely to be that procedural reforms are not needed to reduce the effects of bias on departmental decision-making. The high cost of the failure to institute reforms will be that implicit bias will continue to detrimentally affect the outcomes of the department's decision-making procedures.

Let us consider some of the specific ways that implicit bias influences disciplinary decision-making. One area for which we have some data is the effect of implicit bias on publishing. Numerous studies have shown that implicit bias plays a larger role in our decisions when we are asked to make hasty judgments based on little information (Kahneman 2011; Olberding, Irvin, & Ellis 2014; Valian 1998). Like many decision-making procedures in our discipline, journal editing fits this description. Journal editors must make quick decisions about whether usually non-anonymized articles should be sent out to reviewers (Haslanger 2008). Factors besides submission quality play a role in the editorial decision to send a paper out for external review. A perceived lack of "fit" with the journal can be reason enough for a desk rejection. As in hiring, the nebulous notion of fit is easily misused in journal editing as a justificatory vehicle for dismissing work by members of marginalized groups. When *Behavioral Ecology* moved to a system of review in which the editors do not know the names of the authors, the journal saw a 33% increase in the number of articles they published by women authors (Budden et al. 2008).¹² Some philosophy journals practice triple-anonymous review, like *Noûs*, *Philosophy and Phenomenological Research*, *Pacific Philosophical Quarterly*, *Journal of Philosophy*, and *Mind*. Many more, such as *Analysis*, *Philosophers' Imprint*, *Philosophical Quarterly*, and *Canadian Journal of Philosophy* do not. A Lee and Schunn (2010; 2011) survey of seventeen general philosophy journals found that while 90% of them withheld authors' identities from referees, 81% failed to withhold author identities from editors. All but one of the journals surveyed sometimes used desk rejections, allowing for editor bias to affect editorial decisions.

As in journal editing, hiring is an area in which evaluators must make high-stakes decisions about candidates after spending very little time with their application materials. Search committees are expected to make first cuts to their applicant pools after skimming applicants' CVs and the first pages of their writing samples. While there are few specific data on the impact of implicit bias on hiring in philosophy, studies on hiring in the sciences suggest that implicit bias has a significant impact. Moss-Racusin et al. (2013) found that gender was a major factor in hiring decisions when they sent identical CVs labeled either "John" or "Jennifer" to professors of biology, chemistry, and physics at research universities and asked them to evaluate applicants for a position as lab manager.

12. This suggests that anonymizing processes have the potential to reduce the effects of implicit bias.

The scientists ranked the male applicant significantly higher in competence and hireability and said they were far more willing to mentor the male applicant. Scientists also offered the male applicant a significantly higher starting salary than the female applicant (\$30,238.10 and \$26,507.94, respectively). While there is general evidence of implicit bias in academic hiring, we need more data on how these issues crop up specifically in the context of hiring in philosophy. For this reason, empirically investigating the effects of implicit bias on hiring practices within the discipline must be a priority.

The 2014 APA Feminism and Philosophy Newsletter explores best practices for hiring and identifies a number of important reforms necessary for reducing the effects of implicit bias and stabilizing job criteria that can oscillate depending on the gender, race, sexuality, class, and disability status of the candidate. Prior to the on-campus interview stage these efforts include developing clear job criteria before reading any dossiers and applying these to all candidates as uniformly as possible, screening for potential negative bias triggers and flagging files that contain them so that committee members may attend to them with additional time and care, maintaining a critical awareness of professional privileges accruing based on the Matthew effect,¹³ forgoing or using highly structured interviews, and anonymizing writing samples and distributing them to the entire faculty (Olberding, Irvin, & Ellis 2014). Once a small pool of candidates is chosen to provide job talks, efforts must be made to ensure that the best performance conditions possible are established for each candidate and that uniformity across multiple candidate visits is prioritized. This involves maintaining a standardized format for on-campus interviews, providing candidates with questions in advance, striving to ensure that no candidate is interviewing under solo status, treating information gained during on-site visits as supplementary to that contained in candidates' dossiers, minimizing the influence of information obtained during more informal and social aspects of the visit such as conversations over meals, and eschewing discussions of the notoriously loaded criterion of "fit" with the department.

Many of these reforms serve to systematize and formalize evaluations that are based on complex aggregations of huge quantities of data using differently weighted variables. Accepting the need for such reforms in hiring practices involves recognizing the limits of individuals' abilities to accurately and uniformly weigh a multitude of factors across a number of cases. It also requires an awareness of the wide range of subtle circumstantial factors that can trigger

13. Merton (1968) coined this term in the sociology of science to refer to the positive feedback mechanisms by which already well-known and prestigious scientists gain more credit and recognition for their work than do relatively unknown scientists. This leads to greater resources and productivity, creating a feedback loop of prestige accumulation. For an explanation of how the Matthew effect functions within philosophy graduate programs, see Schliesser (2013).

unconscious implicit biases and cause them to influence our conscious deliberations. An account of rationality that de-emphasizes the ways our reasoning and decision-making processes are affected by situational factors makes it difficult to recognize how decision-making procedures magnify the influence of implicit bias and may thus prevent the necessary reforms from being implemented.

5. Rationality, Objectivity, and Self-Evaluation

Consider what the effects of the idealized rational self are likely to be at the scale of the individual. How might an acceptance of the idealized rational self lead one to overestimate one's own ability to overcome bias? Perhaps those who embrace such a notion take themselves to be closer to achieving the standard than the average person. Or maybe a higher evaluation of one's own rational abilities makes one more likely to embrace a conception of the idealized rational self. In this section, I consider possible relationships among self-evaluations of rationality and objectivity, implicit bias, and the idealized rational self.

In general, people evaluate themselves more positively than they evaluate others. This is known as the better-than-average (BTA) effect, and it is seen in self-evaluations across a wide range of skills and tasks (Guenther & Alicke 2010). Is the BTA effect more pronounced in philosophers than in other academics when it comes to evaluations of rationality? Could the percentage of philosophers who rate themselves as smarter and more rational than average itself be even higher than average? Consider how disciplinary focus on logical consequence and argument structure may lead philosophers to take themselves to be more rational than the average non-philosopher. Antony hypothesizes that "Philosophers . . . broadly regard themselves as smarter than others in the humanities and think of themselves as particularly sensitive to fallacious or ungrounded reasoning" (2012: 236). This itself might be enough to make philosophers skeptical that they are subject to the same implicit biases as everyone else, resulting in an attitude that procedural reforms aimed at correcting for effects of implicit bias are great for everyone else but unnecessary for themselves. If this is right, then "not only will philosophers be intractable about instituting reforms in evaluative practices, they will be more susceptible to the unconscious biases they do have" (Antony 2012: 236). Philosophers' confidence in their own lack of bias and ability to overcome its effects may thus serve to inoculate their biases from critical reflection and other efforts to address and undermine them.

Many practices within philosophy promote a conception of objectivity as something that operates at the scale of individual reasoning. Philosophical training involves developing skills such as identifying and avoiding logical fallacies, providing necessary and sufficient conditions for given concepts, elucidating the

logical consequences of a view, formulating counterexamples to a view in light of its logical consequences, and revising a view repeatedly so that it is impervious to counterexamples. The quality checks used in philosophy tend to be both purely a priori and accomplishable by single individuals. Contrast this with the sort of training that scientists more commonly receive. Many of science's tasks are inherently collaborative and are often not the sort of thing that can be completed by a single individual. These may include tasks like research design, data collection and analysis, running models and simulations, or recognizing patterns in data. In these cases, the quality of the scientific work is not always transparent to individual scientists without the use of computational checks. The collaborative and diffuse nature of the activities that the scientific method comprises are some of the reasons that Longino (1990) argues for a view on which what is recognized as scientific knowledge must be thought of as the output of a social process. She similarly argues objectivity itself ought to be conceived of as something that applies at the scale of social processes rather than at the scale of individual practitioners. Whereas objectivity in science is easily conceived of as an emergent, group-level phenomenon, objectivity in philosophy tends to be more commonly portrayed as a feature of the reasoning of individuals.¹⁴ That norms within philosophy often portray a single person's reasoning as capable of achieving truth on its own likely contributes to philosophers' high levels of confidence in their own judgments.

How might we understand the interplay of the BTA effect with both the demographics and the self-conception of philosophy described above? Brown (2012) showed that the BTA effect increases after participants experience a threat to their feelings of self-worth. Plausibly, when one is faced with a suggestion that procedural reforms are necessary to overcome the effects of bias on one's own decision-making, this is experienced as a threat to one's self-worth.¹⁵ And such a suggestion may be experienced as an even greater threat by someone whose self-conception is rooted primarily in a sense of their own rationality and responsiveness to reasons. Resistance to the pervasiveness of implicit bias and the necessity of procedural reforms to overcome it may serve to function as a defense mechanism against a perceived threat to one's sense of self. This could

14. This is not to say that objectivity in philosophy cannot similarly be conceived of as a property of community-level knowledge practices, just that it does not tend to be. Longino, in fact, holds the same view of objectivity in philosophy that she does for objectivity in science.

15. We know, for instance, that white men experience institutional commitments to diversity as threatening. Pro-diversity organizational messaging makes white men more likely to believe that they are being discriminated against and more likely to believe they are being treated unfairly (Dover, Major, & Kaiser 2016). It is possible that this perception of threat could produce the relevant increase in the BTA effect and thus increase resistance to the claim that one's own implicit biases need to be addressed. Further empirical investigation into how this phenomenon functions within professional philosophy departments is needed.

operate alongside and even magnify the effects of a perceived threat to one's advantaged status. Research on the pervasive effects of implicit bias may create an especially pronounced perceived threat to one's advantaged status when it supports the belief that one may have received their position in part due to biases in their favor rather than by earning it through purely meritocratic means. Thus, a commitment to the idealized rational self might actually be strengthened in the face of evidence showing that bias results in systemic discrimination against marginalized groups. Since stronger levels of belief in the existence of meritocracy result in lowered likelihood of participation in collective action to address systemic discrimination against marginalized groups (Foster, Sloto, & Ruby 2006), this would create a positive feedback loop between commitment to the idealized rational self and resistance to addressing and remedying the effects of implicit bias.

But let's consider an objection. What if philosophers really do receive special training that insulates them from the effects of implicit bias on their reasoning? If this is the case, then philosophers might actually be better than those in other disciplines at recognizing and rooting out bias without the need for institutional and procedural reforms. This is certainly worth investigating empirically even if only to rule it out as a possibility. The fact that philosophers are specifically trained in good reasoning does not make them immune from implicit biases or from irresponsible use of heuristics in decision-making. Expertise can provide false confidence in one's judgments while failing to protect against common biases and mistakes in reasoning. Professionals who are trained as experts in certain areas often exhibit the same sorts of errors in reasoning as non-experts. Consider Kahneman and Tversky's (1972) research on the use of the representativeness heuristic as a common error in statistical reasoning. They found that people tend to judge something as being more probable if it is more "representative" or better reflects the most salient features of its kind even if the base rate of the feature in the population in question is lower than the alternative. Kahneman (2011) offers an anecdote of giving the classic "feminist bank teller" scenario to a statistician colleague.¹⁶ The colleague immediately fell prey to the same fallacious reasoning usually exhibited by non-statisticians, using fit with stereotype to guide his answer rather than obvious assumptions about the base rates of each feature or bundle of features. Or think, for instance, of the wide range of

16. This classic problem is an example of the conjunction fallacy. The problem offers a description of a woman, Linda, who studied philosophy in college, was interested in social justice, and marched for equal rights. The question then posed to subjects asks which scenario is more likely: a) Linda is a bank teller, or b) Linda is a bank teller who is active in the feminist movement. Despite the fact that the structure of the problem makes a more probable, most subjects rate b as more probable because it the option that is more representative of the salient features noted in the description.

emphatically negative responses to Marilyn vos Savant's column in *Parade* magazine on the Monty Hall Problem (Tierney 1991). Vos Savant received around 10,000 letters, the vast majority of which disagreed with her correct answer to the problem. The most vehement criticism came from professional mathematicians and scientists who wrote in to chastise her for her "error" and "explain" to her how obvious it was that there was a fifty percent chance of winning the car regardless of whether one stayed with one's original choice or switched doors. Many of these expert critics demonstrated extraordinary levels of confidence in their own incorrect judgments. It is clear that expertise in an area does not inoculate one against mistakes, and in contrast, can even promote false confidence or epistemic hubris.

Recent work on priming and objectivity in social psychology suggests further countervailing evidence for the hypothesis that philosophers are likely to be especially good at rooting out bias. Uhlmann and Cohen (2007) show that being primed with a sense of one's own objectivity leads to greater reliance on gender bias in mock hiring scenarios. These results suggest that there is even reason to think that the special training that philosophers receive could actually function to increase the effects of implicit bias by promoting self-perceptions of objectivity.

Examining the details of Uhlmann and Cohen's experiments will help to clarify their connections to academic philosophy. The series of experiments used only male subjects, as those who are in positions to hire applicants for high-status jobs are overwhelmingly male. Subjects in the experimental group were primed with a sense of their own objectivity by filling out a questionnaire that asked them to rank on an 11-point scale how strongly they agreed with statements such as, "When forming an opinion, I try to objectively consider all facts I have access to," "My judgments are based on a logical analysis of the facts," and "My decision making is rational and objective." Participants were then provided with a mock hiring scenario in which they were asked to imagine being an executive evaluating a job candidate for a position as factory manager. Each participant received the dossier of a single job candidate. All dossiers were identical except for the candidate's name, which was either "Lisa" or "Gary." Their dossier described them as technically proficient and organized but interpersonally unskilled. Participants rated the strength of the applicant with respect to a series of traits. Those participants who were primed with a sense of personal objectivity rated the female candidate less favorably than the male candidate. Whereas the control group gave approximately equal ratings to the male and female candidates, primed participants rated the male candidate an average of 5.67 out of 10 and the female candidate an average of 3.94.¹⁷ Merely answering

17. Standard deviation of 1.32 and 1.53 respectively.

a few questions about their decision-making abilities was enough to produce a sizeable gender-discrimination effect.

Uhlmann and Cohen hypothesize that the explanation of their results is that decision-makers' sense of personal objectivity acted as a disinhibitor of discrimination by providing them with a justification for trusting their initial impressions, which are the ones likeliest to be influenced by bias and stereotypes.

When people believe that they are objective, they feel licensed to act on biases whose influence they may have otherwise suppressed due to personal and social inhibitions A sense of personal objectivity, we suggest, gives rise to an 'I think it, therefore it's true' mindset: People assume that their thoughts and beliefs are, by virtue of being theirs, valid and therefore worthy of being acted upon. (Uhlmann & Cohen 2007: 208)

Self-perceived objectivity thus acts as a conduit for whatever subconscious stereotypic thoughts and beliefs the decision-maker harbors.

These results suggest that philosophy's disciplinary self-conception as objective, universal, and rational could actually function to magnify the effects of implicit bias on philosophers' reasoning. This would be the case if philosophy's self-conception promotes philosophers' self-perceptions of objectivity in a way similar to priming. The special training that philosophers receive may work in conjunction with the a priori bent of much philosophical methodology to partially explain philosophy's overrepresentation of white men. One possibility is that philosophy's heavy methodological reliance on a priori reasoning as a guide to discovering truth acts as a frequently reinforced primer of a sense of one's own objectivity for those in the discipline. A belief in or use of the model of the idealized rational self may also act as such a primer. Since those in the discipline are already very likely to be white and male and to have biases in favor of those who are also white and male, their well-developed sense of personal objectivity provides them the necessary confidence to trust their quickly formed impressions about candidate qualifications and quality and to make important decisions about such things as hiring without concern that their judgments are influenced by bias in the same way that everyone else's judgments are. This would be another way in which philosophy's disciplinary relationship to implicit bias is somewhat special—among academics, philosophers take themselves to be uniquely insulated from the effects of implicit biases on their reasoning.

Of course, the causal connection between being a philosopher and having a magnified sense of one's objectivity could also work in the other direction. If this were the case, then those who were already highly disposed to take themselves to be objective thinkers would be more attracted than others to the discipline of philosophy. This would overpopulate philosophy with thinkers who have

a highly developed sense of their own objectivity, which could also be enough to tempt them into a subconscious trust in their initial judgments of candidate quality. This would achieve the same effect as priming someone with a sense of their own objectivity. It is worth noting the difference between subjects being temporarily primed to think they are objective and subjects having a stable disposition over time to judge themselves to be objective. Though these are distinct phenomena they may well produce a similar overreliance on one's initially formed impressions and judgments of candidate qualifications and quality.¹⁸

One interesting empirical prediction that comes out of this hypothesis is that philosophy departments with a heavier empirical focus would be pro tanto more likely to institute the relevant procedural reforms to counteract the effects of implicit bias than would departments whose focus leans more strongly toward the rationalist and a priori. Another empirical prediction is that disciplines outside of philosophy that have a greater-than-average reliance on a priori methodologies and employ something close to a conception of the idealized rational self would also have greater overrepresentation of white men. Economics, with its reliance on both methodological individualism and its conception of homo economicus, would seem to fulfill both criteria. Theoretical physics and mathematics both employ a heavy reliance on a priori reasoning. All three of these disciplines are known to have significant overrepresentation of white men.¹⁹

The possible correlation between a priori methodology and disciplinary demographics may also bear an interesting connection to Leslie, Cimpian, Meyer and Freeland's (2015) research. The researchers showed that the level of underrepresentation of women in a field could be predicted by how strongly its practitioners embrace the belief that innate, raw talent or natural brilliance is essential to success in that field. Their findings revealed that philosophy had by far the highest acceptance of field-specific ability beliefs among all the disciplines and some of the lowest number of PhDs who were women, African American, or

18. For reasons to be skeptical of the aggregated notion of candidate quality, see Bright (2015).

19. In the 2014-2015 academic year, roughly a quarter of permanent positions in Canadian economics departments are held by women. 40% of undergraduate economics majors are women. In 2008-2009, only 37% of undergraduate students, 43% of Master's students and 33% of PhD students in economics were women (Canadian Women Economists Network 2015). It is worth emphasizing again that implicit bias is by no means the only or even the primary factor determining the underrepresentation of marginalized groups within these disciplines. In the case of economics, for instance, the underrepresentation of women coincides with a generally hostile discipline. Wu (2017), in a recent study of anonymous online message boards of economics faculty and graduate students, revealed that the 30 words most uniquely associated with discussions of women included sexualized words for body parts and misogynistic and objectifying themes (e.g., tits, anal, feminazi, slut, hot, vagina, boobs, pregnant, cute, gorgeous, dated, horny), while the 30 words most uniquely associated with discussions of men were generally innocuous and relevant to the academic discipline of economics (e.g., advisor, mathematician, pricing, textbook, and Wharton).

Asian American.²⁰ The disciplines that rank more highly on the field-specific-ability belief scale also exhibit a greater reliance on a priori or non-empirical methodologies, and those that employ more empirical methodologies tend to rank lower.²¹ This seeming correlation, like the other hypotheses offered here, would benefit from further empirical investigation.

6. Idealizations, Ideal Theory, and Active Ignorance

Consider a final objection to the hypothesis that the prevalence of the idealized rational self in philosophy functions to protect practices that allow bias to flourish. Many of those who embrace a version of the idealized rational self take it to be merely an idealization. Holding something to be an ideal standard is compatible with believing that the standard is rarely if ever achieved. It does not follow that embracing an account of the idealized rational self commits one to believing that anyone achieves such a standard. Perhaps philosophers who accept a picture of the idealized rational self are insulated from its proposed distorting effects by their recognition that it is merely an idealization. In this section, I explain why acknowledging the idealized rational self to be mere idealization is probably not enough to prevent its hypothesized effects of magnifying biases. I also discuss reasons to expect that the notion of the idealized rational self is a mechanism of active ignorance that functions to obscure structural injustices in a way similar to “color-blind” racial ideology.

I start by turning to Mills’s (2005) work on the uses and limitations of idealizations and ideal theory in addressing systemic injustice. Mills’s thesis about ideal theory as ideology is intimately connected to practices of active ignorance, or the social production of unknowing. As Bailey notes, while ignorance is frequently conceived of as “a gap in understanding” and as something “that can be corrected by an effort to move toward certainty by finding the missing information or running the experiment again,” an important project of feminist epistemology and epistemologies of ignorance more generally has been to identify ways in which “ignorance is often an active social production” (2007: 77). I suggest that the picture of the idealized rational self can function as a mechanism of active ignorance.

Mills (2005) argues that ideal theory functions to obscure important features of the world from our understanding. The role that ideal theory has played in

20. Their methods did not include an analysis of the correlations between field-specific ability beliefs and the underrepresentation of PhDs who are situated at the intersections of multiple systems of oppression, such as African American women or Asian American women.

21. Specifically, the disciplines that rank highest on Leslie’s scale are, in order, Philosophy, Math, Music Composition, Physics, Economics, and Classics.

moral and political philosophy has resulted in moral theory's inability to capture and illuminate the realities of systemic injustice. Moral theory that is characterized by its focus on the ideal has consistently failed to engage with historical actualities that, while non-ideal, are essential to understanding and addressing injustice and oppression in the actual world. The problem lies not in ideal theory's use of idealization, argues Mills, but in its implicit rejection of the relevance of features that fall outside of its idealizations. He writes, "Ideal theory either tacitly represents the actual as a simple deviation from the ideal, not worth theorizing in its own right, or claims that starting from the ideal is at least the best way of realizing it" (Mills 2005: 168). It is thus not ideal theory itself but rather the way that ideal theory is commonly used and engaged that produces active ignorance by masking the systemic injustices that constitute our actual social and political landscape.

Active ignorance arises when the idealizations produced by ideal theory obscure the contingent suboptimal aspects of our reality that are nonetheless integral to understanding the actual nature and function of the system or phenomenon modeled by the idealization. The obscurantist features of ideal theory include the idealization of human capacities, ideal conceptions of social institutions, idealized social ontologies, silence on oppression, and an idealized cognitive sphere. Yap puts the issue as follows,

A world in which no one is subordinated due to gender or race may be 'ideal,' but it is not the world we live in. As such, a moral theory that ignores such subordination ignores some very important facts about the situations in which the theory is to be applied. (2016: 63)

Yap notes that the problems with ideal theory arise not just in moral theory but within epistemology as well. She writes, "An epistemology that ignores the almost aggressive ignorance associated with privilege also ignores some very important facts about the situations in which it is to be applied" (Yap 2016: 63). The failure to consider the effects of social identity on knowledge produces inaccurate and potentially harmful idealizations.

In describing the idealized cognitive sphere, Mills gestures at how his criticisms of ideal theory in moral philosophy can also be applied to theories of rationality:

As a corollary of the general ignoring of oppression, the consequences of oppression for the social cognition of these agents, both the advantaged and the disadvantaged, will typically not be recognized, let alone theorized. A general social transparency will be presumed, with cognitive obstacles minimized as limited to biases of self-interest or the intrinsic dif-

faculties of understanding the world, and little or no attention paid to the distinctive role of hegemonic ideologies and group-specific experience in distorting our perceptions and conceptions of the social order. (2005: 169)

Because the idealized rational self is conceived of as a reasoner who is not affected either by hegemonic ideologies such as white supremacy or by group-specific experience, the idealization offers no tools to correct for these cognitive distortions on human reasoning. The harms of idealization result from erasing actual biases that arise both from the social location of the reasoner and from the stereotypes and controlling images that are applied to those being evaluated. When philosophers make evaluative judgments about candidate excellence, for instance, they may be affected both by motivated reasoning about their own position as well as by unconscious biases and preconceptions about applicants who are members of marginalized groups. Feminists, critical race theorists, and science studies scholars have long argued that our conceptions of rationality have been both explicitly and implicitly restricted in their applications to those who are white and male (Collins 2000; Eze 1997; Gould 1981; Heikes 2010; Longino 2005). Assumptions about whiteness as the default race and maleness as the default gender set the normative standard for who is imagined to be a “typical” cognizer. May writes, “Subjugated knowers are often denied equal cognitive authority, in part because ‘marked’ embodiments are perceived to undermine rationality” (2014: 95). The conception of the idealized rational self ignores these factors in its modeling of human reasoning, just as ideal moral theory ignores them in its treatment of justice.

Ideal moral theory not only functions to obscure the non-ideal nature of human cognition in ways that exacerbate the effects of oppression by rendering them invisible, it also functions to obscure central features of non-ideal social and political reality, like the fact that the social contract is really a racial contract (Mills 1997). The racial contract is the political structure (conceived as an agreement in social contract theory) that prescribes “the differential privileging of the whites as a group with respect to the nonwhites as a group, the exploitation of their bodies, land, and resources, and the denial of equal socioeconomic opportunities to them” (Mills 1997: 11). Moral theory, for Mills, functions to cover over and distract from the facts of white supremacy. He writes,

The frustrating problem that nonwhites have always had, and continue to have, with mainstream political theory, is not with abstraction itself . . . but with an idealizing abstraction that abstracts away from the crucial realities of the racial polity. The shift to the hypothetical, ideal contract encourages and facilitates this abstraction, since the eminently nonideal features of the real world are not part of the apparatus. (Mills 1997: 76)

As Mills points out for the idealizing abstraction of the social contract, there is no “part of the apparatus” that allows us to identify that race is fundamental to structuring moral and political life, both in the United States and globally following European colonialism. While the social contract was once explicitly racialized in the days of formalized *de jure* white supremacy it now hides its nature as a racialized agreement and instead purports to be a set of inherently race-less assumptions (Mills 1997: 73).

The persistence of the racial contract depends on this continued invisibility of the structure of white supremacy. It is essential to its functioning that white supremacy maintains a widespread (among whites) denial of its existence. The denial of the power structure of white supremacy and the denial of racialized experiences form the two elements of what is known as color-blind racial ideology (CBRI), a modern form of racism that is researched in psychology and sociology.²² Following Frankenberg (1993), Neville, Awad, Brooks, Flores, and Bluemel (2013) break down CBRI into its dual components of power-evasion and color-evasion. Color-evasion refers to the “denial of racial differences by emphasizing sameness” and power-evasion refers to the “denial of racism by emphasizing equal opportunities” (Neville et al. 2013: 455). Both of these dimensions of CBRI obscure and protect racial inequality. Neville et al. write,

CBRI is a dominant racial ideology or worldview that serves to justify and explain away racial inequalities in society; it is thus one type of ideology that is used ‘to [help] sustain the social hierarchy while maintaining a perspective that provides the cover of innocence’ (APA Presidential Task Force on Preventing Discrimination and Promoting Diversity 2012, 4) (Neville et al. 2013: 458).

Color-evasion and power-evasion are both essential to the persistence of white supremacy as a foundational organizing principle of contemporary political, economic, and social institutions. They are important components of white ignorance, the term Mills uses to describe the form of active ignorance that arises from the cognitive and epistemological norms mandated by the racial contract of white supremacy. Specifically, white supremacy prescribes and demands “white misunderstanding, misrepresentation, evasion, and self-deception on matters related to race” (Mills 1997: 18). The pretense of racelessness makes the

22. For a discussion of the ableism inherent to terms like ‘colorblindness’ and ‘colorblind racism’ that metaphorically equate blindness with ignorance, see (Annamma, Jackson, & Morrison 2017). I acknowledge the need to move away from linguistic practices that tacitly uphold structures and ideologies of ableism. I thus follow Grzanka and Morrison (2017) in embracing the language of ‘color-evasive racism.’ I also note that the invocation of evasion in ‘color-evasive racism’ more accurately reflects the active nature of the ignorance at the core of the phenomenon.

racialized structure of the contract invisible to the vast majority of those who are not directly acquainted with it via their experiences of being racialized as non-white. Mills explains that this mandated ignorance of white supremacy entails that there is “no conceptual point-of-entry to start talking about the fundamental way in which (as all nonwhites know) race structures one’s life and affects one’s life chances” (1997: 76). It is difficult to convince others of the harms and need to resist something whose existence they refuse to acknowledge in the first place.

Like the non-racialized conception of the social contract, the idealized rational self is an idealizing abstraction that, while once explicitly racialized and gendered, now purports to be genderless and raceless.²³ This move, like the move from *de jure* white supremacy to color-evasive racism, is a mechanism of active ignorance. It allows theorists of the idealized rational self to treat “the present as a somehow neutral baseline” (Mills 1997: 76) from which rationality can be theorized with no reference to the ideological legacies of racism, sexism, and colonialism that continue to govern our deployment of the concept. Like the social contract that claims to be race-less in a world structured by white supremacy, the idealized rational self provides no “conceptual point-of-entry” to identify ways that race and gender are fundamental to our actual judgments about what thoughts and evaluations are rational and who is capable of producing them. That the idealized rational self abstracts away from embodied and identifying features altogether results in a practice that upholds active ignorance and obscures important features of reality. The idealized rational self picture thus functions as a tool of both color-evasion and power-evasion.

The move from treating the idealized rational self as a normative ideal to treating the present as a “somehow neutral baseline” is a conceptual sleight of hand that moves from a purportedly normative context to a descriptive one without acknowledging that such a shift has taken place. These equivocations serve to obfuscate the existence of bias and its effects on reasoning.²⁴ The unacknowledged shift in domain also parallels a common type of color-evasive

23. It is often the case that claiming that something is neutral or apolitical is itself a non-neutral, political move. As Mills acknowledges, “Toni Morrison points out that there are contexts in which claiming racelessness is itself a racial act” (1997: 76).

24. To illustrate the sort of slide from the normative to the descriptive that I have in mind, imagine a philosopher on a hiring committee who offers the following reasoning for why the committee need not make procedural changes to the way they evaluate applicants: “As a Kantian, I think people are rational agents. Rationality involves being able to evaluate things impartially. Since I’m a rational agent, I should be able to evaluate these applications impartially without anonymizing them. So I don’t think we need to anonymize applications.” Note that the “should” here functions not just as a normative claim that the philosopher ought to be able to evaluate applications in an unbiased way but as a descriptive prediction that they will actually be able to. In this line of reasoning, the philosopher begins with a normative conception of rationality and of what a rational agent is and concludes through equivocation that self-oversight is all they need to keep their biases at bay. Though this slip between the normative and descriptive is certainly

racism, in which an inference is made from the presumption that an ideal world would be a world without race to the conclusion that race is not real, or that race ought to be treated as nonexistent starting immediately (e.g., “Race and/or skin color have been a source of unjust discrimination. Therefore, I don’t see color.”). If one fails or simply refuses to recognize that the actual world is disparately and unjustly structured, then they will not be willing or able to engage in efforts to transform the unjust structure. It is no surprise, for instance, that a belief in meritocracy is a strong predictor of opposition to affirmative action programs (Son Hing, Bobocel, & Zanna 2002). If there is no problem in the first place, there is no need to fix it. Indeed, attempting to “fix” a non-existent injustice might itself produce an injustice. The often-hidden assumption present in pictures of the idealized rational self is that no injustice follows from our conception of rationality or from our tendency to deploy such a conception in harmfully disparate ways. This assumption conceals the need to address the actual injustices that our uneven deployment of the concept of rationality produces and maintains.

The idealized rational self provides a justification for ignoring the non-ideal aspects of cognitive agents’ own purportedly rational judgments. That the idealized rational self is seen as a legitimate philosophical picture, even if not the dominant one, provides the “cover of innocence” for failing to engage with these non-ideal features of reality, such as the fact that philosophical conceptions of rationality have tended to exclude certain groups of people from their domains. The impact of this active ignorance is not harmless. Research on color-evasive racism shows a growing body of evidence that establishes a connection between higher levels of color-evasion and power-evasion and greater demonstration of modern racist attitudes (Awad, Cokley, & Ravitch 2005). Among whites, attempting to avoid acknowledging the race of others, for instance, results in increased demonstrations of racial intolerance (Apfelbaum, Pauker, Ambady, Sommers, & Norton 2008; Norton, Sommers, Apfelbaum, Pura, & Ariely 2006). This is one example of how failing to acknowledge and engage with the unjust, racialized nature of social reality actually leads to greater injustice.

Like practices of color-evasion and power-evasion, focusing on the idealized rational self involves turning away from the non-ideal, racialized, gendered dimensions of human cognition. The conception of the idealized rational self masks the actual injustice produced by members of dominant groups who overestimate their own neutrality while routinely assessing marginalized groups and their judgments as less rational. Idealizations that abstract away from power asymmetries and do not provide explicit acknowledgement of this abstraction or of the potential harms that can follow from it engage in the pretense that the

not limited to conceptions of the idealized rational self, its application to it can have particularly insidious effects.

present provides a “somehow neutral baseline.” The idealized rational self portrays the landscape of rationality as a politically neutral place when in fact it is not. Implicitly gendered and racialized conceptions of rationality that masquerade as neutral and apolitical are used as weapons against marginalized groups. The idealized rational self shields those who wield it from criticism by preserving an illusion of neutrality.

While the criticisms of the idealized rational self offered here are in part normative, they also provide further opportunity for empirical investigation. Possible lines of research might involve having subjects read and rate their levels of agreement with a description of the idealized rational self and then complete a task that involves i) a mock hiring scenario of the Uhlmann and Cohen variety, ii) a survey rating levels of support for implementing the sorts of best practices described in Section 3, iii) a survey evaluating participants’ levels of modern racism (or other aversive attitudes toward marginalized groups), or iv) a survey in which subjects estimate current levels of discrimination based on race or gender (as well as other dimensions of marginalization such as sexuality, class, or disability). The statement(s) read by the experimental group could simply be descriptive, portraying the idealized rational self picture as an accurate picture of human cognition. Or the statements could make normative claims, such as “people ought to be able to overcome the effects of their physical state on their reasoning” or “people should be able to transcend cultural norms when thinking rationally.” This latter option would get at the potentially greater issue raised that the idealized rational self picture leads to the effects hypothesized even when it is explicitly presented as a normative ideal. These experimental conditions would be contrasted with the results of subjects in the control group who complete the same tasks only after reading and rating their agreement with statements unrelated to rationality. This research would offer confirming or disconfirming evidence for the concern that the idealized rational self functions to obscure both non-ideal aspects of human reasoning and current levels of structural inequality; it would also provide a starting point to empirically investigate the claim that acceptance of the idealized rational self conception makes proponents less likely to support measures necessary to reduce the influence of bias on decision-making procedures.

7. Concluding Remarks

I have identified a number of features of the conception of the idealized rational self, and I have suggested that certain of these are in tension with our best understanding of implicit bias. I have hypothesized their potential links to a disciplinary-wide resistance to acknowledging and addressing the problematic

role of implicit bias in institutional decision-making procedures. Features common to philosophical models of rationality likely create obstacles to understanding and addressing the role of bias in human reasoning. The idealized rational self can produce distorting effects on models and theories of knowledge. Treating such models as mere idealizations will not necessarily preclude them from producing harmful impacts.

There is more work to be done here. The hypotheses in this paper would benefit significantly from further empirical investigation, and I aim to have shown that they warrant such inquiry. It must be a priority of the discipline to investigate if and how our unique conceptual and methodological assumptions produce a special resistance to taking bias seriously. Depending on the empirical findings, reform in philosophy may need to be focused at the methodological center of the discipline rather than only at its margins. But we don't have to hold off until certain causal connections are established before we begin improving admissions and hiring procedures guided by the best practices outlined in Olberding, Irvin, and Ellis (2014) and the APA Good Practices Guide (APA Task Force on A Best Practices Guide 2017). Benétreau-Dupin and Beaulac (2015) argue that we need not wait for more data about the underrepresentation of women in philosophy before we begin our efforts to increase their representation. With respect to marginalized groups within philosophy more generally, I wholeheartedly agree.

References

- American Psychological Association Commission on Ethnic Minority Recruitment, Retention, and Training in Psychology (1996). *How to Recruit and Hire Ethnic Minority Faculty*. Retrieved from <http://www.apa.org/pi/oema/resources/brochures/how-to.aspx>
- Amodio, David M. and Patricia G. Devine. (2006). Stereotyping and Evaluation in Implicit Race Bias: Evidence for Independent Constructs and Unique Effects on Behavior. *Journal of Personality and Social Psychology*, 91(4), 652–661
- Annamma, Subini Ancy, Darrell D. Jackson, and Deb Morrison (2017). Conceptualizing Color-Evasiveness: Using Dis/ability Critical Race Theory to Expand a Color-Blind Racial Ideology in Education and Society. *Race Ethnicity and Education*, 20(2), 142–162.
- Anscombe, Gertrude (1958) *Modern Moral Philosophy*. *Philosophy*, 33(124), 1–19.
- Antony, Louise (2012). Different Voices or Perfect Storm: Why Are There So Few Women in Philosophy? *Journal of Social Philosophy*, 43(3), 227–255.
- APA Task Force on a Best Practices Guide (2017). *Good Practices Guide*. American Philosophical Association. Retrieved from http://c.ymcdn.com/sites/www.apaonline.org/resource/resmgr/docs/Good_Practices_Guide.pdf
- Apfelbaum, Evan P., Kristin Pauker, Nalini Ambady, Samuel R. Sommers, and Michael I.

- Norton (2008). Learning (Not) to Talk about Race: When Older Children Underperform in Social Categorization. *Developmental Psychology*, 44(5), 1513–1518. <https://doi.org/10.1037/a0012835>
- Awad, Germaine H., Kevin Cokley, and Joseph Ravitch (2005). Attitudes toward Affirmative Action: A Comparison of Color-Blind versus Modern Racist Attitudes. *Journal of Applied Social Psychology*, 35(7), 1384–1399. <https://doi.org/10.1111/j.1559-1816.2005.tb02175.x>
- AWIS-MAA Joint Task Force on Prizes and Awards (2012). Avoiding Implicit Bias: Guidelines for MAA Selection Committees. Retrieved from https://www.maa.org/sites/default/files/pdf/sections_archived/Sections_ImplicitBiasStatement.pdf
- Bach, Kent (1997). Do Belief Reports Report Beliefs? *Pacific Philosophical Quarterly*, 78(3), 215–241.
- Bailey, Alison (2007). Strategic Ignorance. In Shannon Sullivan and Nancy Tuana (Eds.), *Race and Epistemologies of Ignorance* (77–94). State University of New York Press.
- Bealer, George (2000). A Theory of the A Priori. *Pacific Philosophical Quarterly*, 81(1), 1–30.
- Beattie, Geoffrey, Doron Cohen, and Laura McGuire (2013). An Exploration of Possible Unconscious Ethnic Biases in Higher Education: The Role of Implicit Attitudes on Selection for University Posts. *Semiotica*, 2013(197), 171–201.
- Beebee, Helen (2013). Women and Deviance in Philosophy. In Katrina Hutchison and Fiona Jenkins (Eds.), *Women in Philosophy: What Needs to Change?* (61–84). Oxford University Press.
- Benétreau-Dupin, Yann and Guillaume Beaulac (2015). Fair Numbers: What Data Can and Cannot Tell Us About the Underrepresentation of Women in Philosophy. *Ergo*, 2(3), 59–81.
- Bengson, John (2015). The Intellectual Given. *Mind*, 124(495), 707–760.
- Bodenhausen, Galen (1990). Stereotypes as Judgmental Heuristics: Evidence of Circadian Variations in Discrimination. *Psychological Science*, 1(5), 319–322
- Botts, Tina F., Liam K. Bright, Myisha Cherry, Guntur Mallarangeng, and Quayshawn Spencer (2014). What is the State of Blacks in Philosophy? *Critical Philosophy of Race*, 2(2), 224–242.
- Braun, David (1998). Understanding Belief Reports. *The Philosophical Review*, 107(4), 555–595.
- Braun, David (2006). Illogical, But Rational. *Noûs*, 40(2), 376–379.
- Bright, Liam (2015). Against Candidate Quality. Manuscript in preparation.
- Brown, Jonathon (2012). Understanding the Better than Average Effect: Motives (Still) Matter. *Personality and Social Psychology Bulletin*, 38(2), 209–219.
- Budden, Amber, Tom Tregenza, Lonnie W. Aarssen, Julia Koricheva, Roosa Leimu, and Christopher J. Lortie. (2008) Double-Blind Review Favours Increased Representation of Female Authors. *Trends in Ecology & Evolution*, 23(1), 4–6.
- Canadian Women Economists Network (2015). CWEN/RFE Report on the Status of Women in Canadian Economics, 2015. Retrieved from <http://www.cwen-rfe.org/wp-content/uploads/2015/09/CWENRFE-report-2015.pdf>
- Carruthers, Peter (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford University Press.
- Ceci, Stephen J., Donna K. Ginther, Shulamit Kahn, and Wendy M. Williams (2014). Women in Academic Science: A Changing Landscape. *Psychological Science* 15(3), 75–141.

- Cherry, Myisha and Eric Schwitzgebel (2016, March 4) Like the Oscars, #PhilosophySo-White. Los Angeles Times. Retrieved from <http://www.latimes.com/opinion/op-ed/la-oe-0306-schwitzgebel-cherry-philosophy-so-white-20160306-story.html>
- Collins, Patricia Hill (2000). *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. Routledge.
- Cooley, Erin, Keith B. Payne, Chris Loersch, and Ryan Lei (2015). Who Owns Implicit Attitudes? Testing a Metacognitive Perspective. *Personality and Social Psychology Bulletin*, 41(1), 103–115.
- Crimmins, Mark (1992). *Talk about Belief*. MIT Press.
- Crouch, Margaret A. and Lisa H. Schwartzman (2012). Introduction. *Journal of Social Philosophy*, 43(3), 205–211.
- Cullison, Andrew (2007). Privileged Access, Externalism, and Ways of Believing. *Philosophical Studies*, 136(3), 305–318.
- Descartes, Rene (1996) *Meditations on First Philosophy*. John Cottingham (Ed.). Cambridge Texts in the History of Philosophy. Cambridge University Press.
- Dotson, Kristie (2012). How is this Paper Philosophy? *Comparative Philosophy*, 3(1), 3–29.
- Dover, Tessa, Brenda Major, and Cheryl Kaiser (2016). Members of High-Status Groups are Threatened by Pro-Diversity Organizational Messages. *Journal of Experimental Social Psychology*, 62, 58–67.
- Dovidio, John F. and Samuel L. Gaertner (2000) Aversive Racism and Selection Decisions. *Psychological Science*, 11(4), 315–319.
- Dutt, Kuheli, Danielle L. Pfaff, Ariel F. Bernstein, Joseph S. Dillard, and Caryn J. Block (2016). Gender Differences in Recommendation Letters for Postdoctoral Fellowships in Geoscience. *Nature Geoscience*, 9, 805–808.
- Engelbert, Mark and Peter Carruthers (2010). Introspection. *WIREs Cognitive Science*, 1, 245–253. <https://doi.org/10.1002/wcs.4>
- Eze, Emmanuel Chukwudi (1997). The Color of Reason: The Idea of ‘Race’ in Kant’s Anthropology. In Emmanuel Chukwudi Eze (Ed.), *Postcolonial African Philosophy: A Critical Reader* (103–131). Blackwell.
- Feit, Neil (2001). Rationality and Puzzling Beliefs. *Philosophy and Phenomenological Research*, 63(1), 29–55.
- Fernández, Jordi (2003). Privileged Access Naturalized. *Philosophical Quarterly*, 53(212), 352–372.
- Fiske, Susan (1998). Stereotyping, Prejudice and Discrimination. In D. T. Gilbert, S.T. Fiske, and G. Lindzey (Eds.), *Handbook of Social Psychology* (357–411). McGraw-Hill.
- Forbes, Graeme (1990). The Indispensability of Sinn. *Philosophical Review*, 99(4), 535–563.
- Foster Mindi D., Lisa Sloto, and Richard Ruby (2006). Responding to Discrimination as a Function of Meritocracy Beliefs and Personal Experiences: Testing the Model of Shattered Assumptions. *Group Processes and Intergroup Relations*, 9(3), 401–411.
- Frankenberg, Ruth. (1993). *White Women, Race Matters: The Social Construction of Whiteness*. University of Minnesota Press.
- Gehrman, Kristina (2016, February). Ethics and the Great Chain of Being. In Spring Symposium: Ethics and Humanity’s Place in Nature. Symposium conducted at the University of Tennessee.

- Gehrman, Kristina (2017). Ethics and the Great Chain of Being. Manuscript submitted for publication.
- Good, Catherine, Aneeta Rattan, and Carol S. Dweck (2012). Why Do Women Opt Out? Sense of Belonging and Women's Representation in Mathematics. *Journal of Personal and Social Psychology*, 102(4), 700–717.
- Gould, Stephen Jay (1981). *The Mismeasure of Man*. W.W. Norton and Company.
- Green, Alexander R., Dana R. Carney, Daniel J. Pallin, Long H. Ngo, Kristal L. Raymond, Lisa I. Iezzoni, and Mahzarin R. Banaji (2007) Implicit Bias among Physicians and Its Prediction of Thrombolysis Decisions for Black and White Patients. *Journal of General Internal Medicine*, 22(9), 1231–1238.
- Grzanka, Patrick and Daniel R. Morrison (2017). Postracial Fantasies and the Reproduction of Scientific Racism. *The American Journal of Bioethics*, 17(9), 65–67. <https://doi.org/10.1080/15265161.2017.1353179>
- Guenther, Corey L. and Mark D. Alicke, (2010). Deconstructing the Better than Average Effect. *Journal of Personality and Social Psychology*, 99(5), 755–770.
- Haslanger, Sally (2008). Changing the Ideology and Culture of Philosophy: Not by Reason (Alone). *Hypatia*, 23(2), 210–222.
- Haslanger, Sally (2013, September 2). Women in Philosophy? Do the Math. *New York Times*. Retrieved from <https://opinionator.blogs.nytimes.com/2013/09/02/women-in-philosophy-do-the-math/>
- Healy, Kieran (2011, February 4) Gender Divides in Philosophy and Other Disciplines [Web log post]. Retrieved from <http://kieranhealy.org/blog/archives/2011/02/04/gender-divides-in-philosophy-and-other-disciplines/>
- Heikes, Deborah K. (2010). Rationality and Feminist Philosophy. *Continuum*.
- Hermanson, Sean (2017). Implicit Bias, Stereotype Threat, and Political Correctness in Philosophy. *Philosophies*, 2(12), 1–17.
- Huemer, Michael (2007). Compassionate Phenomenal Conservatism. *Philosophy and Phenomenological Research*, 74(1), 30–55.
- Jaschik, Scott (2011, March 30) A Call to Shun. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/news/2011/03/30/philosophers_consider_what_to_do_about_sexual_harassment
- Jennings, Carolyn Dicey, Patrice Cobb, and David W. Vinson (2016). Academic Placement Data and Analysis: An Update with a Focus on Gender [Web log post]. *Blog of the APA*. Retrieved from <http://blog.apaonline.org/2016/05/03/academic-placement-data-and-analysis-an-update-with-a-focus-on-gender/>
- Jost, John T., Laurie A. Rudman, Irene V. Blair, Dana R. Carney, Nilanjana Dasgupta, Jack Glaser, and Curtis D. Hardin (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.
- Kahneman, Daniel (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.
- Kahneman, Daniel and Amos Tversky (1972). Subjective Probability: A Judgment of Representativeness. *Cognitive Psychology*, 3(3), 430–454.
- Kelly, Daniel and Erica Roedder (2008). Racial Cognition and the Ethics of Implicit Bias. *Philosophy Compass*, 3(3), 522–540.
- Korman, Daniel Z. (2005). Law Necessitarianism and the Importance of Being Intuitive. *Philosophical Quarterly*, 55(221), 649–657.

- Kripke, Saul (1979). A Puzzle about Belief. In Avishai Margalit (Ed.), *Meaning and Use* (239–283). D. Reidel.
- LaCroix, Patricia P. (1985) Sex in Recs. *The Journal of College Admission*, Fall, 24–26.
- Lee, Carole J. and Christian D. Schunn (2010). Philosophy Journal Practices and Opportunities for Bias. *American Philosophy Association Newsletter on Feminism and Philosophy*, 10(1), 5–10.
- Lee, Carole J. and Christian D. Schunn (2011). Social Biases and Solution for Procedural Objectivity. *Hypatia*, 26(2), 352–373.
- Leiter, Brian (2014, Dec 12). Implicit Bias, Stereotype Threat, and the Empirical Literature [Web log post]. *Leiter Reports: A Philosophy Blog*. Retrieved from <http://leiterreports.typepad.com/blog/2014/12/implicit-bias-stereotype.html>
- Leiter, Brian (2016, May 5). Recent PhD Placement Data: Women More Likely than Men to Secure Permanent Positions within Two Years of the PhD [Web log post]. *Leiter Reports: A Philosophy Blog*. Retrieved from <http://leiterreports.typepad.com/blog/2016/05/recent-phd-placement-data-women-more-likely-than-men-to-secure-permanent-positions-within-two-years-.html>
- Leslie, Sarah-Jane, Andrei Cimpian, Meredith Meyer, and Edward Freeland (2015). Expectations of Brilliance Underlie Gender Distributions across Academic Disciplines. *Science*, 347(6219), 262–265.
- Lewis, David (1981). What Puzzling Pierre Does Not Believe. *Australian Journal of Philosophy*, 59(3), 283–289.
- Longino, Helen (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press.
- Longino, Helen (2005). Circles of Reason: Some Feminist Reflections on Reason and Rationality. *Episteme*, 2(1), 79–88.
- Madera, Juan M., Michelle R. Hebl, and Randi C. Martin (2009). Gender and Letters of Recommendation for Academia: Agentic and Communal Differences. *Journal of Applied Psychology*, 94(6), 1591–1599. <https://doi.org/10.1037/a0016539>
- Madva, Alex (2017). Implicit Bias, Moods, and Moral Responsibility. *Pacific Philosophical Quarterly*. Advance online publication. <https://doi.org/10.1111/papq.12212>
- Marcus, Ruth Barcan (1981). A Proposed Solution to a Puzzle about Belief. *Midwest Studies in Philosophy*, 6(1), 501–510.
- Markie, Peter J. (2013). Rational Intuition and Understanding. *Philosophical Studies*, 163(1), 271–290.
- May, Vivian M. (2014). “Speaking into the Void”? Intersectionality Critiques and Epistemic Backlash. *Hypatia*, 29(1), 94–112.
- McConnell, A. R. and J. M. Leibold (2001). Relations among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes. *Journal of Experimental Social Psychology*, 37(5), 435–442.
- McCullagh, Mark (2002). Self-Knowledge Failures and First Person Authority. *Philosophy and Phenomenological Research*, 64(2), 365–380.
- McGlone, Michael (2009). Understanding Kripke’s Puzzles about Belief. *Philosophy Compass*, 4(3), 487–514.
- McPherson, Lionel K. (2011, December 12). On Black Underrepresentation and Progress in the Profession [Web log post]. *New APPS*. Retrieved from <http://www.newapps-blog.com/2011/12/new-apps-on-black-underrepresentation-and-progress-in-the-profession.html#moree>

- Merton, Robert K. (1968). The Matthew Effect in Science. *Science*, 159(3810), 56–63.
- Meyers, Diana (2010). Feminist Perspectives on the Self. In Edward N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Spring 2010 ed.). Retrieved from <https://plato.stanford.edu/archives/spr2010/entries/feminism-self/>
- Mills, Charles (1997). *The Racial Contract*. Cornell University Press.
- Mills, Charles (2005). 'Ideal Theory' as Ideology. *Hypatia*, 20(3), 165–183.
- Mills, Charles (2007) White Ignorance. In Shannon Sullivan and Nancy Tuana (Eds.), *Race and Epistemologies of Ignorance* (11–38). State University of New York Press.
- Milkman, K. L., M. Akinola, M. and D. Chugh (2015, April 13). What Happens Before? A Field Experiment Exploring How Pay and Representation Differentially Shape Bias on the Pathway Into Organizations. *Journal of Applied Psychology*. Advance online publication. <https://doi.org/10.1037/ap10000022>
- Morgan, W.B., K. B. Elder, and E. B. King (2013). Emergence and Reduction of Bias in Letters of Recommendation. *Journal of Applied Social Psychology*, 43(11), 2297–2306.
- Moss-Racusin, Corinne A., John F. Dovidio, Victoria L. Brescoll, Mark J. Graham, and Jo Handelsman. (2013). Science Faculty's Subtle Gender Biases Favor Male Students. *Proceedings of the National Academy of Sciences*, 109(41), 16474–16479.
- Neville, Helen A., Germinie H. Awad, James E. Brooks, Michelle P. Flores, and Jamie Bluemel. (2013). Color-Blind Racial Ideology: Theory, Training, and Measurement Implications in Psychology. *American Psychologist*, 68(6), 455–466.
- Norton, Michael I., Samuel R. Sommers, Evan P. Apfelbaum, Natassia Pura, and Dan Ariely (2006). Color Blindness and Interracial Interaction. *Psychological Science*, 17(11), 949–953. <https://doi.org/10.1111/j.1467-9280.2006.01810.x>
- Olberding, Amy, Sherri Irvin, and Stephen Ellis (2014) Best Practices for Fostering Diversity in Tenure-Track Searches. *American Philosophy Association Newsletter on Feminism and Philosophy*, 13(2), 27–36. Retrieved from http://c.ymcdn.com/sites/www.apaonline.org/resource/resmgr/feminism_newsletter/feminismv13n2.pdf
- Paxton, Molly, Carrie Figdor, and Valerie Tiberius (2012). Quantifying the Gender Gap: An Empirical Study of the Underrepresentation of Women in Philosophy. *Hypatia*, 27(4), 949–957.
- Petty, Richard E., Gary L. Wells, Timothy C. Brock (1976). Distraction Can Enhance or Reduce Yielding to Propaganda: Thought Disruption versus Effort Justification. *Journal of Personality and Social Psychology*, 34(5), 874–884.
- Raja, Sheela and Joseph P. Stokes (1998). Assessing Attitudes toward Lesbians and Gay Men: The Modern Homophobia Scale. *Journal of Gay, Lesbian, and Bisexual Identity*, 3(2), 113–134.
- Reskin, Barbara (2000) The Proximate Causes of Employment Discrimination. *Contemporary Sociology*, 29(2), 319–328.
- Richard, Mark (1990). *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them*. Cambridge University Press.
- Ross David A., Dowin Boatright, Marcella Nunez-Smith, Ayana Jordan, Adam Chekroud, and Edward Z. Moore (2017) Differences in Words Used to Describe Racial and Gender Groups in Medical Student Performance Evaluations. *PLoS ONE*, 12(8), e0181659. <https://doi.org/10.1371/journal.pone.0181659>
- Ruíz, Elena Flores (2014). Musing: Spectral Phenomenologies: Dwelling Poetically in Professional Philosophy. *Hypatia*, 29(1), 196–204.
- Saul, Jennifer (2011, March 28). Why Is Philosophy So White, Male, and Middle Class?

- Implicit Bias, Stereotype Threat and Under-Represented Groups in Philosophy. Presentation conducted at St. George's Church, Sheffield, UK.
- Saul, Jennifer (2013, August 15) Philosophy Has a Sexual Harassment Problem. Salon. Retrieved from https://www.salon.com/2013/08/15/philosophy_has_a_sexual_harassment_problem/
- Saul, Jennifer (2015, January 23). Women in Philosophy. The Philosopher's Magazine. Retrieved from <https://www.philosophersmag.com/opinion/9-women-in-philosophy>
- Scanlon, Tim M. (1998). *What We Owe to Each Other*. Belknap Press of Harvard University Press.
- Schliesser, Eric (2013, September 3) On Boy-Wonders in Philosophy [Web log post]. New APPS: Art, Politics, Philosophy, Science. Retrieved from <http://www.newappsblog.com/2013/09/on-boy-wonders-in-philosophy.html>
- Schmader, T., J. Whitehead, and V. H. Wysocki (2007). A Linguistic Comparison of Letters of Recommendation for Male and Female Chemistry and Biochemistry Job Applicants. *Sex Roles*, 57(7–8), 509–514. <https://doi.org/10.1007/s11199-007-9291-4>
- Schuman, Rebecca (2014, February 3) Nasty and Brutish: A Scandal in Colorado Reveals that Bullying Bros Still Plague University Philosophy Departments. Slate. Retrieved from http://www.slate.com/articles/life/education/2014/02/sexual_harassment_in_philosophy_departments_university_of_colorado_boulder.html
- Sekaquaptewa, Denise and Mischa Thompson (2003). Solo Status, Stereotype Threat, and Performance Expectancies: Their Effects on Women's Performance. *Journal of Experimental Social Psychology*, 39(1), 68–74.
- Sesardic, Neven, Rafael De Clercq (2014, December 10). Women in Philosophy: Problems with the Discrimination Hypothesis. National Association of Scholars. Retrieved from: https://www.nas.org/articles/women_in_philosophy_problems_with_the_discrimination_hypothesis
- Son Hing, Leanne S., Ramona D. Bobocel, and Mark P. Zanna (2002). Meritocracy and Opposition to Affirmative Action: Making Concessions in the Face of Discrimination. *Journal of Personality and Social Psychology*, 83(3), 493–509.
- Sosa, David (1996). The Import of the Puzzle about Belief. *Philosophical Review*, 105(3), 373–402.
- Sosa, Ernest (2007). Experimental Philosophy and Philosophical Intuition. *Philosophical Studies*, 132(1), 99–107.
- Staats, Cheryl (2014). State of the Science: Implicit Bias Review 2014. Kirwan Institute.
- Stalnaker, Robert (1981). Indexical Belief. *Synthese*, 49(1), 129–151.
- Stalnaker, Robert (1987). Semantics for Belief. *Philosophical Topics*, 15(1), 177–190.
- Sue, Derald Wing, Christina M. Capodilupo, Aisha M. B. Holder (2008). Racial Microaggressions in the Life Experience of Black Americans. *Professional Psychology: Research and Practice*, 39(3), 329–336.
- Suinn, Richard, Norma Abeles, Patricia Arredondo, Beverly Greene, Douglas C. Halderman, Zari Hedayat-Diba, . . . and Joseph L. White (2005). APA Presidential Task Force on Enhancing Diversity Final Report. American Psychological Association Office of Ethnic Minority Affairs. Retrieved from <https://www.apa.org/pi/oema/resources/taskforce-report.pdf>
- Symons, John (2008). Intuition and Philosophical Methodology. *Axiomathes*, 18(1), 67–89.

- Taschek, William (1998). On Ascribing Beliefs: Content in Context. *The Journal of Philosophy*, 95(7), 323–353.
- Tierney, John (1991, July 22). Behind Monty Hall's Doors: Puzzle, Debate, and Answer? *New York Times*. Retrieved from <http://www.nytimes.com/1991/07/21/us/behind-monty-hall-s-doors-puzzle-debate-and-answer.html?pagewanted=all&mcubz=0>
- Trix, Frances and Carolyn Psenka (2003). Exploring the Color of Glass: Letters of Recommendation for Female and Male Medical Faculty. *Discourse and Society*, 14(2), 191–220.
- Valian, Virginia (1998). *Why So Slow? The Advancement of Women*. MIT Press.
- Uhlmann, Eric Luis and Geoffrey L. Cohen (2007). "I Think it, Therefore it's True": Effects of Self-Perceived Objectivity on Hiring Discrimination. *Organizational Behavior and Human Decision Processes*, 104(2), 207–223.
- Wu, Alice (2017). *Gender Stereotyping in Academia: Evidence from Economics Job Market Rumors Forum* (Senior thesis). Princeton University. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3051462
- Yap, Audrey (2016). Feminist Radical Empiricism, Values, and Evidence. *Hypatia*, 31(1), 58–73.