Implicit Bias, Unconscious Discrimination, and the Nature of Philosophical Inquiry

Lieke Asma

1. Introduction

Research suggests that academic philosophy has an exceptionally hard time becoming more diverse. In contrast to STEM disciplines for example, the number of women receiving PhDs in philosophy has not gone up in the period from 1990 until 2015, but remained at around thirty percent.[1] Research from 2017 and 2018 shows that women compose at most 25 percent of U.S. philosophy faculty.[2] What is more, even considering this low number, citations and article submissions from women to top ranked journals are low.[3]

These data require an explanation. Why do women still not play a substantial role in philosophical inquiry? Several scholars have pointed out that the problem is not independent from the nature of philosophy.[4] Louise Antony distinguishes between two models: Different Voices and Perfect Storm. While some scholars adopt Different Voices and propose that gender differences explain why women leave philosophy, for example because women tend to have opposed intuitions[5] or dislike of the

1 Holtzman, Geoffrey S.: Rejecting beliefs, or rejecting believers? On the importance and exclusion of women in philosophy, in Hypatia 31 (2016) 2, 293–312, 301. Leslie, Sarah-Jane et al: Expectations of brilliance underlie gender distributions across academic disciplines, in: *Science*, 347 (2015) 6219, 262–265.

2 Schwitzgebel, Eric/Jennings, Carolyn Dicey: Women in philosophy. Quantitative analyses of specialization, prevalence, visibility, and generational change, in: Public Affairs Quarterly 31 (2017) 2, 83–105

3 Wilhelm, Isaac/Conklin, Sherry Lynn/Hassoun, Nicole: New data on the representation of women in philosophy journals: 2004–2015, in Philosophical Studies 175 (2018), 1441–1464.

4 For example, Antony, Louise: Different voices or perfect storm. Why are there so few women in philosophy?, in Journal of Social Philosophy 43 (2012) 3, 227–255. Holtzman: Rejecting beliefs.

5 Buckwalter, Wesley/Stich. Stephen: Gender and philosophical intuition, in: Knobe, Joshua/Nichols, Shaun (eds.): Experimental Philosophy: Volume 2, New York 2014, 307–346.

combative nature of philosophy[6], Antony argues for the Perfect Storm view. In philosophy, she maintains, certain problematic forces converge, interact, and intensify each other. For one thing, even though women may not intrinsically dislike the combative manner of philosophy, it clashes with accepted gender norms for women.[7] It puts women in a double bind: if they respect gender norms, they are likely to be dismissed intellectually, and if they don't, they risk being seen as rude or domineering.[8] Another factor is that in philosophy the credibility of the researcher may play a more important role compared to other fields of inquiry[9], because philosophical inquiry is less firmly anchored by facts or data independent from the researcher, and depends more on personal intuitions.[10] Additionally, Leslie *et al.* argue that it is often assumed that philosophy requires innate brilliance, a factor that is sensitive to gender-coded stereotypes as well.[11]

Several scholars have pointed out that implicit biases may be an important contributor to this perfect storm.[12] Of course, explicit sexism and racism occur[13], but the problem also is that people unconsciously associate philosophy and rationality with white males. Conversely, women (and persons of color) are typically associated with everything philosophy is not: emotionality, subjectivity, and the body.[14] Such gender schemas, as Sally Haslanger and Antony take from Virginia Valian, condition our perceptions, shape our normative expectations, and influence the way in

6 For example: Beebee, Helen: Women and deviance in philosophy, in: Hutchison, Katrina/Jenkins, Fiona (eds.): Women in Philosophy. What Needs to Change? Oxford 2013, 61–80.

7 Antony: Different voices, 238. See also: Friedman, Marilyn: Why should we care?, in: Hutchison, Katrina/Jenkins, Fiona (eds.): Women in Philosophy. What Needs to Change? Oxford 2013, 21–38.

8 Antony: Different voices, 238. Haslanger, Sally: Changing the ideology and culture of philosophy. Not by reason (alone), in: Hypatia 23 (2008) 2, 210–223.

9 Kirloskar-Steinbach, Monika: Diversifying philosophy. The art of non-domination, in: Educational Philosophy and Theory 51 (2019) 14, 1490–1503.

10 For example: Hutchison, Katrina: Sages and cranks. The difficulty of identifying first-rate philosophers, in: Hutchison, Katrina/Jenkins, Fiona (eds.): Women in Philosophy. What Needs to Change? Oxford 2013, 103–126.

11 Leslie et al: Expectations of brilliance. Saul, Jennifer: Implicit bias, stereotype threat, and women in philosophy, in: Hutchison, Katrina/Jenkins, Fiona (eds.): Women in Philosophy. What Needs to Change? Oxford 2013, 54.

12 Antony: Different voices; Beebee: Women and deviance in philosophy; Haslanger, Changing the ideology; Leuschner, Anna: Why So Low? On Indirect Effects of Gender Bias in Philosophy, in: Metaphilosophy 50 (2019) 3, 231–249; Saul: Implicit bias, stereotype threat. For a critical perspective, see Thompson, Morgan: Explanations of the gender gap in philosophy, in: Philosophy Compass 12 (2017) 3, e12406.

13 See, for example: https://beingawomaninphilosophy.wordpress.com/ (11/9/2023).

14 Haslanger, Changing the ideology.

which we search for information and evaluate it, for example our decision to read certain work and whether we judge it as insightful or worthwhile.[15]

In this picture of implicit bias, the assumption is that unconscious psychological states like gender schemas causally explain why women remain underrepresented in philosophy. These psychological states can account for the gap between what people explicitly believe about women philosophers, and how they actually treat and evaluate them and their work.[16] However, we have to pay attention to the fact that these psychological explanations are invoked to explain a peculiar kind of discrimination as well. After all, we do not point to unconscious psychological states if we want to explain why a professor hires only male PhD candidates, if he explicitly states that only men have the innate brilliance to be successful philosophers. What is more, implicit attitudes may result in conscious evaluations. My expectation that my male partner will pay for dinner, for example, may be caused by gender schemas I am not conscious of, but I still consciously expect my partner to pay.[17] That means that in order to have a full picture of the problem, we should pay attention to the nature of unconscious discrimination itself, and examine how this might be related to the nature of philosophy.

In this chapter, I will argue that philosophy is particularly susceptible to unconscious discrimination, because its nature interacts with mechanisms that contribute to discrimination remaining hidden from view. In section 2, I provide two explanations of unconscious discrimination: misattribution and justification. In section 3, I argue that these mechanisms are likely to play a role in philosophy and, therefore, that unconscious discrimination is likely to occur. Section 4 concludes the paper.

2. The Nature and Explanation of Unconscious Discrimination

Before I can continue, a definition of unconscious discrimination is required. Discrimination, understood as unfair treatment, typically involves responding to facts that are irrelevant for the evaluation, for example considering the person's gender or ethnicity when evaluating the quality of a paper or presentation. Accordingly, I take discrimination to be unconscious if the person is not conscious of such irrelevant facts playing a role in their evaluation.[18]

15 Antony: Different voices; Haslanger, Changing the ideology; Valian, Virginia: Why so slow? The advancement of women, Cambridge 1999.

16 Beebee: Women and deviance in philosophy; Saul: Implicit bias, stereotype threat, 45.

17 See Asma, Lieke J. F.: Implicit bias as unintentional discrimination, in: Synthese, 202 (2023), 129. Gawronski, Bertram/Hofmann, Wilhelm/Wilbur, Christopher J: Are "implicit" attitudes unconscious?, in: Consciousness and cognition 15 (2006) 3, 485–499.

18 Asma: Implicit bias.

But how can we fail to realize which facts we are responding to? In the remainder of this section, I provide two explanations: misattribution and justification.

Misattribution involves failing to accurately identify the cause of your evaluation.[19] It comprises three elements: (1) the true cause of the feeling or evaluation, (2) the apparent cause, and (3) mistaking of one cause for the other.[20] In one classic example, men misattribute their arousal, that was in fact caused by a precarious bridge, to the attractiveness of the woman directly across it.[21] Misattribution is a well-known phenomenon in psychology, and as scholars in the field have emphasized, in the complexities of daily life, having a hard time attributing feelings or arousal to a certain source seems quite common. We do not always know what brought about our feelings, and we may very well ascribe it to a source that did not actually cause it.

Research shows that misattribution typically takes place under certain conditions. We do not misattribute a positive feeling resulting from having lunch with a dear friend to the movie we saw the night before, or our anger towards a dangerous driver to the old lady walking by. Factors that contribute to misattribution are: (1) the events taking place or facts being presented to the agent in close proximity in time and place,[22] (2) ambiguity of the target in relation to the aim,[23] and (3) applicability or conceptual overlap of the internal experience with the target or goal.[24] Simply put: misattribution can occur when the facts, evaluation, and goal allow persons to fill in the blanks in different ways, and when it makes sense to think that the apparent cause did cause the evaluation. An attractive woman can indeed cause arousal in a (heterosexual) man, for example. Another factor that may play a role, but has not been explicitly discussed in relation to misattribution, is the extent to which the ex-

19 March, David S./Olson, Michael A./Fazio, Russell H.: The implicit misattribution model of evaluative conditioning, in: Social Psychological Bulletin, 13 (2018) 3, 1–25. Payne, B. Keith et al: An inkblot for attitudes: affect misattribution as implicit measurement, in: Journal of Personality and Social Psychology 89 (2005) 3, 277–293.

20 Payne, B. Keith et al.: A process model of affect misattribution, in: Personality and Social Psychology Bulletin 36 (2010) 10, 1397–1408.

21 Dutton, Donald G./Aron, Arthur P.: Some evidence for heightened sexual attraction under conditions of high anxiety, in: Journal of Personality and Social Psychology 23 (1974), 510–517.

22 Jones, Christopher R./Fazio, Russell H./Olson, Michael A: Implicit misattribution as a mechanism underlying evaluative conditioning, in: Journal of Personality and Social Psychology 96 (2009) 5, 933–948. March et al: implicit misattribution model. Payne et al: Inkblot for attitudes. Payne et al: Process model.

23 See Payne et al: Inkblot for attitudes.

24 Ecker, Yael/Bar-Anan, Yoav: Applicability increases the effect of misattribution on judgment, in: Cognition and Emotion 33 (2019) 4, 709–721. Ecker, Yael/Bar-Anan, Yoav: Conceptual overlap between stimuli increases misattribution of internal experience, in: Journal of Experimental Social Psychology 83 (2019), 1–10.

planation we can give is in line with our self-image and folk psychology.[25] Am I the kind of person that would experience arousal while crossing a precarious bridge, or does it make more sense to think the attractive woman caused it?

This brings me to justification. Often, if we do not respond to the facts correctly, we recognize making a mistake. Sometimes the mistake is obvious, for example when you try to open the door with the bike key or realize that you are about to cut an apple while making spaghetti. In other cases, we recognize the mistake when we have time to reflect. This is clear from an influential experiment in the field of implicit bias: the shooter task. In this task, participants have to decide quickly whether to 'shoot' a black or white man depending on whether the object they are holding is a gun or a benign object.[26] Even though participants tend to 'shoot' black men even if they hold a benign object, they recognize their mistake and are able to correct for their bias if they have the time to do so.[27]

But having time to reflect does not lead to success in all cases, for example in Uhlmann and Cohen's study.[28] Their participants had to choose a new police chief. They either had to choose between a streetwise Michelle and formally educated Michael, or between a formally educated Michelle and a streetwise Michael. The results showed that most participants selected Michael, and that they justified their decision by weighing the credentials of the Michael and Michelle differently: whichever credential Michael possessed, they took that to be more important. This result can be explained in terms of misattribution: even if the name 'Michael' caused the positive evaluation of that candidate, this could be misattributed to whichever credential he possessed. But what is more, even though these participants had time to reflect on their decision, they still did not recognize that they were using irrelevant information. They did not correct themselves like the participants in the shooter task did. This, I maintain, is because they could justify whichever decision they made: choosing the male candidate for a job is not inherently sexist and, what is more, streetwiseness and formal education are both important for police chiefs. From the perspective of the participants, regardless of the condition they were in, they could justify their decision and did not perceive themselves as making a mistake. Conversely, shooting a black man holding a benign object cannot be justified; it is obviously wrong.

25 Stafford, Tom: The perspectival shift: how experiments on unconscious processing don't justify the claims made for them, in: Frontiers in Psychology 5 (2014), 1067.

26 For example: Payne, B. Keith/Correll, Joshua: Race, weapons, and the perception of threat, in: Advances in Experimental Social Psychology 62 (2020), 1–50.

27 Payne & Correll: Race, weapons.

28 Uhlmann, Eric Luis/Cohen, Geoffrey L.: Constructed criteria. Redefining merit to justify discrimination, in: Psychological Science 16 (2005) 6, 474–480.

Taken together, misattribution and justification can explain how discrimination can be unconscious in itself, not simply because it is caused by an unconscious psychological state. When we misattribute the source of our evaluation, i.e., when we are mistaken about the fact we are responding to, *and* we can justify our evaluation, it is possible that we respond to irrelevant facts without being conscious of doing so, i.e., without realizing that we are discriminating. Crucially, that means that the characteristics of the situation in which we evaluate play a crucial role. Does it allow us to fill in the blanks in a way that the evaluation does not involve unfair treatment? Does it offer alternative explanations that would justify the evaluation?

3. Unconscious Discrimination and the Nature of Philosophical Inquiry

Evaluating philosophical work, I maintain, is one of those situations in which our evaluations can be explained in different ways, because of which misattribution and justification are likely to occur.

A good way to make this clear is by comparing philosophical inquiry to a different case of discrimination: three black women working at NASA in the 1960s, who are portrayed in the book and movie *Hidden Figures*. These women were faced with several kinds of discrimination: their contributions were structurally underestimated or dismissed, they had to use separate toilets and coffeepots, and they were unable to become full-fledged engineers, because engineering could only be studied at a white school for example. Holroyd and Puddifoot have argued that these injustices may be the result of implicit bias.[29]

We should recognize, however, that many of the discriminatory decisions these women were confronted with are cases of *explicit* discrimination. Not allowing a black woman into engineering school because it is only open for white people or urging her to use a separate toilet because of her ethnicity are conscious and intentional responses to irrelevant facts. That is one reason why, near the end of the story, it became obvious that these women were discriminated against. But what is more, people at NASA came to recognize that these women were doing valuable work, and contributed substantially to NASA's success. An important reason for this is that the criteria for success in this context are quite clear: calculations are wrong or right, designs function or not, and the rocket reaches the moon or not. Because of that, there is little room for misattribution or justification. If the calculations a black woman makes are right, dismissing her work must have involved responding to the wrong reasons; her gender or ethnicity for example.

29 Holroyd, Jules/Puddifoot, Katherine: Epistemic injustice and implicit bias, in: Beeghly, Erin/Madva, Alex (eds.): Introduction to Implicit Bias, New York/London 2020, 116–133.

In philosophy, this is different. As other scholars have emphasized, philosophy is more indirectly grounded in facts, and involves intuitions and perspectives of the individual philosopher. What is more, philosophy is more or less characterized by disagreement.[30] This disagreement is not limited to the conclusions drawn, but is also about which method to use[31], or even what philosophers should be aiming for in their work.[32] Moreover, Antony emphasizes that in philosophy the introduction of novel ideas is common, ideas which others may not immediately understand. As a result, "the experience of hearing a novel good idea may – at least initially – be qualitatively identical to the experience of hearing a confused idea."[33] In other words, compared to engineering, in philosophy the criteria for success are less straightforward. We can judge a paper or philosopher as good for a variety of reasons, weigh these reasons in different ways, and sometimes we may not even know which criteria to use or how to use them.

As a consequence, there is ample room for misattribution and justification. The more criteria and ends are open to filling in the blanks, the more we can attribute our evaluation to a source that did not actually cause it but would justify our evaluation. A paper, presentation, or idea leaves a good or bad impression, and there are several explanations that present themselves at more or less the same time which may have contributed to your evaluation: the characteristics of the philosopher, the argumentation, the style of philosophizing, the tone, the examples used, but also the gender or ethnicity of the person. What is more, it is likely that philosophers see themselves as particularly objective and unbiased.[34] Because of this self-image, they may interpret their evaluations as objective and unbiased as well, even if they are not: e.g., "it could not have been gender or ethnicity that influenced my evaluation, there must be something wrong with the argumentation." In other words, even if your evaluation is (partly) caused by irrelevant factors like ethnicity or gender, the nature of philosophical inquiry makes it so that you may not recognize this. Misattribution and justification are lurking, and discrimination can occur and remain hidden from view.

Importantly, the same story may hold for credibility. Several scholars have pointed out that individual credibility plays a more important role in philosophy than in other fields of inquiry, because, indeed, it is more difficult to judge the quality of philosophical work.[35] Of course, credibility may justify our judgments to

30 Bourget, David/Chalmers, David: Philosophers on philosophy. The 2020 philpapers survey, in: Philosophers' Imprint 23 (2023) 1.

31 Ibid., 11.

32 Dotson, Kristie: How is this paper philosophy?, in: Comparative Philosophy 3 (2012) 1, 3–29.

33 Antony: Different voices, 239.

34 Ibid., 236; Erden, Yasemin J.: Identity and bias in philosophy. What philosophers can learn from stem subjects, in: Think 20 (2021) 59, 120.

35 Antony: Different voices, 239; Holtzman: Rejecting beliefs, 302.

some extent; it makes sense to listen to the engineer instead of a physical therapist to evaluate which rocket design is best. But the problem is that in philosophy, our judgment of credibility may be based on irrelevant facts like gender or ethnicity as well, and because of the lack of clear criteria, we may not recognize that our judgment is the result of these facts.[36] Also here, misattribution and justification can explain how we can fail to realize that we use irrelevant criteria to judge a person's credibility as a philosopher.

Finally, these two mechanisms may also explain why women in philosophy themselves could make decisions that maintain the status quo. Recent research shows that differences in publications in top journals is not the result of less manuscripts by women being accepted. Rather, women submit less manuscripts to such journals.[37] Misattribution and justification could explain why women evaluate their papers as not good enough for top journals, think they have good grounds for this judgment, but in fact unintentionally and unconsciously discriminated against themselves. They may think of themselves as not possessing the brilliance to publish in top journals for example[38], but fail to realize that they are taking their gender into account to make this judgment. What is more, by not submitting to top journals, they substantiate facts they think may justify their judgment: "I haven't published in a top journal so far, why would I think I can do so now?".

4. Conclusion, and One Important Implication

In this chapter, I argued, by appealing to misattribution and justification, that philosophy is particularly susceptible to unconscious discrimination. It is relatively likely to believe your evaluation is the result of a relevant fact and justified, even if in reality it is based on irrelevant information like gender or ethnicity. This alternative perspective on the relationship between philosophy and implicit bias supports Antony's Perfect Storm model: philosophical inquiry interacts with psychological mechanisms that contribute to unconscious discrimination.

What is more, the picture I painted emphasizes the particular problematic nature of unconscious discrimination. When we discriminate unconsciously, we use irrelevant facts, but, for example because of misattribution and justification, we do not receive feedback that we made a mistake. All seems fine from the perspective of the person doing the evaluating. Which facts you actually responded to only becomes clear when you compare cases, like in experimental studies such as Uhlmann

36 Hutchison, Sages and cranks.

37 Leuschner: Why so low?

38 Maranges, Heather et al.: Brilliance beliefs, not mindsets, explain inverse gender gaps in psychology and philosophy, in: Sex Roles: A Journal of Research 89 (2023), 801–817.

and Cohen's. That means that in real life it is often difficult to establish whether we have a case of unconscious discrimination. Would the evaluation of the exact same paper or presentation have been different if the person was white or male? That is hard to say.

This puts examples in the literature in a different light. Holroyd, Scaife, and Stafford give the example of Professor P who explicitly strives for fair treatment, but whose "evaluations of the equally good work of black students is slightly less glowing than that of white students."[39] But this is not a fact that can simply be perceived. How should the black students find out that their work was equally good, and that they were discriminated against? Maybe their papers were not as good as they thought. As a result, unconscious discrimination is the ultimate breeding ground for epistemic injustices that we often also do not immediately recognize: Professor P has ample opportunity to dismiss her students' perspective, and argue that even though the black students received lower grades, these grades are justified.[40,41]

39 Holroyd, Jules/Scaife, Robin/Stafford, Tom: Responsibility for implicit bias, in: Philosophy Compass 12 (2017) 3, e12410.

40 See Collins, Patricia Hill: Black feminist thought. Knowledge, Consciousness, and the Politics of Empowerment, New York 2000. Fricker, Miranda: Epistemic injustice. Power and the Ethics of Knowing, Oxford 2007.

41 This work was funded by the Deutsche Forschungsgemeinschaft – AS 667/1-1. An earlier version of this chapter was given at the workshop 'Knowledge, Participation and Power of Discourse' at the Munich School of Philosophy in 2022. I would like to thank the audience for their questions and comments, and Lena Schützle for organizing the workshop and this anthology. Also, I would like to thank Anupam Yadav for her helpful feedback on this chapter.