The Nature of Bias in Decision-Making

Lieke Asma, Munich School of Philosophy

In discussions on psychological bias, typically the focus is on the nature of the psychological state – an attitude or belief – that supposedly causes biased behavior and decision making (e.g., Brownstein, 2018; Greenwald & Banaji, 1995; Holroyd, 2012; Mandelbaum, 2016). Similarly, discussions about algorithmic bias generally take bias to causally explain problematic results: biases have a certain *impact* (e.g., Danks & London, 2017) or lead to problematic *outputs* (Kordzadeh, & Ghasemaghaei, 2021). This approach is limited, however, because it overlooks that in first instance, bias is recognized in problematic results. Only after we've established that a certain decision or pattern of decisions is problematic, we ascribe a certain psychological state to the agent or conclude that the system is flawed. Consequently, we first have to reflect on the sense in which decision or patterns of decisions are biased *in themselves.*

My starting point for reflecting on bias in decision-making is De Houwer's (2019) recent account. He maintains that biased decisions are influenced by cues that are indicative of the social group to which the person belongs. I will argue that this understanding of the nature of bias is (too) broad: it follows that I am also biased if I step of the sidewalk to make room for a person in a wheelchair. Even though we could of course claim that this is a biased response, it deviates from how we normally use the term 'bias'. My response is fully in line with the facts, and does not involve an inference from the information I have to a problematic conclusion, which seems to be the key problem of bias: that we assume mental disability on the basis of physical disability, or leadership qualities on the basis of gender. There is a difference between stepping of the sidewalk for a person in a wheelchair and selecting the

male candidate for a typically masculine job, and an account of bias should be able to account for this distinction.

As an alternative, I will use Antony's (2016) account of bias in belief. Antony maintains that biases are a response to underdetermination: the facts are consistent with an infinite number of theories, and we have to reduce hypothesis space through non-evidential ways (Antony, 2016, p. 161). Bias, then, involves going beyond the facts, but this is not necessarily problematic and in fact unavoidable. We cannot be fully objective. We can, however, evaluate and justify our biases by assessing whether they contribute to finding the truth, i.e., whether they are vindicated and/or ecologically valid (Antony, 2016, pp. 176-177, pp. 183-185).

Similar to bias in belief, I will argue that bias in decision-making is not necessarily problematic and typically unavoidable. Mostly, the evidence does not fully determine which decision is best. When we have to select a new police chief, for example, and have to choose between a candidate that scores higher on streetwiseness and a candidate that scores higher on formal education (see Uhlmann & Cohen, 2005), we have to make inferences and assumptions about what the respective credentials says about how the candidate would function as police chief. We have to fill in the gaps. The goal, then, should not be to be completely objective, but to use those markers that are in fact relevant for the decision at hand. Uhlmann and Cohen's (2005) study shows that many subjects did not use the credential but the candidate's gender to choose the new police chief, i.e., they used irrelevant markers. That is why their decision is biased in a problematic sense.

Taking this approach to bias in decision-making, two important differences with bias in belief surface. Firstly, what counts as a good decision depends on your aim, and whether something counts as a good or problematic bias therefore also depends on your aim. Whether

a person is in a wheelchair is relevant for whether you should make room, but not for deciding whether to ask for directions, for example. Secondly, even if a bias is ecologically valid, e.g., men are generally physically stronger than women, good decision-making requires investigating relevant information about the person(s) you are concerned with. If you want to know who is stronger, you should investigate their strength, not their gender. This means that regardless of whether the bias is ecologically valid, the agent's decision making could still problematic, if the information is irrelevant in relation to that in light of which the agent decides. Bad biases, then, not only lead to injustice, but also to bad decisions (cf. Antony, 2016, p. 185).

These insights, I will finally argue, have important implications for how to think about algorithmic injustice. Generally put, it follows that algorithms cannot simply take over: we continuously have to reflect on our, possibly conflicting, aims and how they relate to the markers we use.

References

Antony, L.M. 2016. Bias: Friend or foe? Reflections on Saulish skepticism. In *Implicit Bias and Philosophy, Volume 1: Metaphysics and Epistemology*, ed. M. Brownstein and J. Saul, 157-190. Oxford: Oxford University Press.

Brownstein, M. 2018. *The Implicit Mind. Cognitive Architecture, the Self, and Ethics*. Oxford University Press.

Danks, D., & London, A. J. (2017). Algorithmic Bias in Autonomous Systems. *IJCAI,17*, 4691-4697.

De Houwer, J. 2019. Implicit bias is behavior: A functional-cognitive perspective on implicit bias. *Perspectives on Psychological Science*.

https://doi.org/10.1177/1745691619855638

Greenwald, A.G. and M.R. Banaji. 1995. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review*. https://doi.org/10.1037/0033-295X.102.1.4

Holroyd, J. 2012. Responsibility for implicit bias. *Journal of Social Philosophy.* https://doi.org/10.1111/j.1467-9833.2012.01565.x

Kordzadeh, N. & Ghasemaghaei, M. (2021). Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems*. https://doi.org/10.1080/0960085X.2021.1927212

Mandelbaum, E. 2016. Attitude, inference, association: On the propositional structure of implicit bias. *Noûs*. https://doi.org/10.1111/nous.12089

Uhlmann, E.L. and G.L. Cohen. 2005. Constructed criteria: Redefining merit to justify discrimination. *Psychological Science.* https://doi.org/10.1111/j.0956-7976.2005.01559.x