Philosophy in Review

Thomas Kelly, "Bias: A Philosophical Study"

Jyoti Kishore

Volume 44, Number 3, August 2024

URI: https://id.erudit.org/iderudit/1113570ar
DOI: https://doi.org/10.7202/1113570ar

See table of contents

Publisher(s)

University of Victoria

ISSN

1206-5269 (print)
1920-8936 (digital)

Explore this journal

Cite this review

Kishore, J. (2024). Review of [Thomas Kelly, "Bias: A Philosophical Study"]. *Philosophy in Review*, *44*(3), 19–21. https://doi.org/10.7202/1113570ar

Thomas Kelly. *Bias: A Philosophical Study*. Oxford University Press 2022. 288 pp. $43.99 USD (Hardcover ISBN); $00.00 USD (Paperback ISBN).

We encounter bias in everyday life among people, objects, or processes. As a phenomenon, bias is commonly perceived as a deviation from its natural course (2). While it is quite common to use the term 'bias' with a negative connotation, some biases are desirable, like inductive biases. They play an indispensable role in cultural evolution and knowledge acquisition. What makes the study of bias even more interesting is that sometimes it is introduced to mitigate another problematic bias. For instance, while training a healthcare allocation algorithm, an algorithmic bias can be introduced in the training data to ensure that the results align with moral standards of fairness to mitigate existing racial bias in society (Fazelpour and Danks, 2021). The multifaceted nature of bias undoubtedly renders it an interesting topic of inquiry across various domains. However, bias has attracted comparatively less attention in philosophy than it has in other disciplines like psychology or statistics.

Thomas Kelly's book *Bias: A Philosophical Study* fills this gap by probing the multilayered nature of bias. He explores a wide range of philosophically exciting theoretical issues about the phenomenon. The book is divided into three parts. The first part offers conceptual fundamentals about bias. The second and third parts explore its relationship with norms and knowledge. One of the book's main contributions is a theoretical framework for thinking about the phenomenon. Calling it the norm theoretic account of bias, Kelly argues for conceptualising bias as typically involving 'a systematic departure from a norm or standard of correctness' (53). He considers these norms to be genuine as opposed to conventional or statistical. Consider the example of an Uber driver in the US who accepts bookings from all but black women. Each time he receives a booking request from a female of African descent, he cancels the request on some pretext. The driver deviates from a genuine moral norm; one should not discriminate based on colour, gender or religious beliefs. He demonstrates a systematic pattern in his deviations, which is refusing to accept booking requests consistently only against black women. This kind of systematic departure from a genuine norm can invoke a charge of bias in a pejorative sense.

In Chapter 1, Kelly acknowledges that there is no exhaustive list of biased things. Diverse things are predicated of bias, like people in social roles, inanimate objects, samples, testimonies, algorithms, etc. He explains the structural features of bias, such as being biased in favour or against something or both, exhibiting directionality. Something can be biased under one circumstance but may not be under another viewpoint, demonstrating relativity. Things with representational content, like data, can also be biased, known as biased representation. There could be cases when parts are biased, but the whole is unbiased; the whole is biased, but none of the parts are biased. While discussing the relative nature of bias, he gives an interesting example of a judge considered biased in one social role but unbiased as a witness. The judge is convinced that the defendant is guilty because they witnessed the accused committing the crime. Kelly asserts that the judge 'doesn't qualify as an unbiased juror' (17). However, his treatment of the judge's role is simplistic and narrow. The judge faces a conflict of interest in his multiple social roles as a juror and a

witness. Kelly does not acknowledge the possibility that the judge could be open to revising his beliefs based on the evidence. They could follow the court proceedings with impeccable objectivity and arrive at the verdict solely based on the evidence. If the evidence contradicts the judge's initial belief, he could set aside his prior conviction and base his judgment solely on the evidence presented, adhering to epistemic norms. This criticism does not undermine the relative nature of bias but asserts that not all conflicts of interest, especially in social roles, can be treated as such.

Kelly argues for robust pluralism about bias in Chapter 2. This means that despite diverse things being biased, none of them is fundamental. There is no hierarchy among biased things. He asserts that people are not the fundamental carriers of bias. When people are biased, it depends on other things related to them, like a biased belief. Hence, people being biased is a derivative matter. A biased process does not necessarily deliver a biased outcome; it can deliver an unbiased outcome and vice-versa. After exploring various salient characteristics of bias in Part I, Kelly explores the central framework and its relationship with norms in Part II.

Kelly offers his fundamental idea of norm theoretic account of framework to understand bias in Chapter 3. The framework is particularly illuminating in illustrating instances in which an accusation of bias invites a countercharge of bias. The perspectival character of bias provides valuable insights as to why it is rational for each party to attribute a charge of bias in cases of systematic disagreements (62). Interestingly, there could be circumstances when rationality and morality require one to be biased. Consequently, sometimes, to comply with a salient norm requires one to depart from another salient norm. Kelly enumerates three approaches to this: liberal, relativist, and priority views.

In Chapter 4, it has been argued that introspection is not only unreliable but a biased method for detecting one's biases. We are likely to consider ourselves unbiased rather than biased while introspecting. It contributes to a bias blind spot, which generates more false negatives than false positives. Kelly emphasises the norm theoretic approach's superiority in explaining bias blind spot over the conventional explanations psychologists provide in terms of naïve realism. The latter explanation assumes that our views are unbiased while those who hold opposing views are biased. The standard account provided by psychologists is unsatisfactory and insufficiently illuminating. On the contrary, the perspectival account is more effective in explaining bias blind spots and biases of introspection.

Kelly develops the concept of biased people in Chapter 5, as many epistemic and prudential norms apply to believers. He explicates how being biased presupposes another failing that the agent is guilty of with respect to a normative claim. This failing is distinct from bias itself. When the judge delivers a biased verdict, he has violated an important norm of impartiality. The same mistake can be committed by an unbiased juror as well. However, what invites the charge of bias is a disposition to commit the mistake systematically. Kelly distinguishes between a liar telling false things and an agent who is not a liar doing the same thing. Both deviate from the norm of truth-telling and commit a moral failure. However, only the liar is subjected to criticism and blame while the other person who unintentionally misleads his audience is not. The difference lies in the fact that the liar misleads his audience deliberately and is morally responsible. The latter commits a

mistake unintentionally and is morally non-culpable. There is an inconsistency with respect to the norm theoretic account in how he uses the liar example. He does not state the systematic pattern in which the liar deviates from the norm of truth-telling. Even if he deliberately deviates from the norm of truth-telling, he deviates randomly as he lies to all his audience. There is no systematic pattern in his deviations. He is like an unbiased jerk (5) who fails to respectfully treat others irrespective of their sex, race, etc.; the liar lies to all his audience. Kelly seems to vacillate regarding the norm theoretic account, pointing to its structural weakness.

Further, in chapters 6 and 7, Kelly explores various norms of objectivity about biases. He expounds upon questions like whether there are systematic deviations from norms without involving biases or biases without involving departure from any genuine norms. Interestingly, deviations from norms of objectivity even once result in biases. It is irrelevant to ask if the departure is systematic or random. Kelly states that paradigmatic cases of bias involve symmetry violations, unlike unbiased cases, which preserve symmetry. So, when the charge of bias in a pejorative sense is in order, it is 'a special case of a more general phenomena' (213) in which an agent systematically departs from a genuine norm. Consequently, it can be grasped that the norm theoretic account is not a reductive analysis of bias. It does not provide necessary and sufficient conditions for attributing bias to something. Kelly clarifies (57) that no strong identity thesis exists between the phenomena and systematic departures from the genuine norm.

The book's last part, entitled *Bias and Knowledge*, comprises three chapters. This part deliberates upon the relationship between bias, knowledge, and scepticism. Kelly explores intriguing questions like, can a biased person know something? Does biased belief qualify as knowledge? Interestingly, one can know something to be true even if one is biased with respect to that knowledge. However, such a true belief, which is also a manifestation of bias, does not qualify as knowledge. This section will be particularly interesting to epistemologists.

In Chapter 11, Kelly summarizes the main themes and conclusions into 125 points. It helps the reader to take stock of valuable points discussed in each chapter.

This book is a remarkable exploration, and an engaging philosophical **examination** of bias supported by ample examples. Despite minor inconsistencies, Kelly's norm-theoretic account is efficient in understanding the central phenomenon and our practices of attributing it. Scholars from various disciplines and emerging domains like algorithmic bias can benefit from this book.

Jyoti Kishore, Indian Institute of Technology, Kanpur