

**Review of Thomas Kelly's *Bias: A Philosophical Study*.
Oxford: Oxford University Press, 2022, x + 288 pp.**

LENNART B. ACKERMANS

Erasmus University Rotterdam

Thomas Kelly's *Bias* is a wide-ranging philosophical study of the concept of 'bias' in the pejorative sense as well as (in one chapter) the non-pejorative sense. The English word 'bias' is used in many different ways to indicate that a person, procedure, or outcome is flawed. The book is concerned with all of these ways. It is an ambitious project looking to give a general theory that explains what bias is and how its different uses relate to each other. It also explores some independent epistemological questions related to the phenomenon of bias.

The book is divided into three parts. Part 1 is called *Conceptual Fundamentals*. It explores some basic questions about the concept of bias, such as to which types of things the label applies and whether some of them are more fundamental. Among other things, bias can be attributed to people (judges, scientists), mental states (reasoning processes), procedures, and outcomes. Kelly argues that there is no fundamental hierarchy among these different types of bias, although sometimes a bias in one sense exists because of bias in another sense—such as when a person is biased because their reasoning process is biased.

Part 2 is called *Bias and Norms*, and it lays out Kelly's theory of bias called the *norm-theoretic account of bias*. In a nutshell, the theory holds that a bias involves a systematic departure from a genuine norm or standard of correctness. There is a great variety of norms that Kelly considers genuine norms: among them are ethical norms, epistemic norms, and norms of rationality. People or things are biased when they depart from such norms systematically, that is, consistently in a particular direction.

For example, suppose a judge generally rules in accordance with the evidence. When the defendant is black, however, non-evidential considerations—namely, prejudice—affect how she comes to her legal decisions. According to the norm-theoretic account of bias, this judge is biased in multiple ways. She is biased with respect to the epistemic norm of believing in accordance with the evidence since her departures from this norm happen systemically (when the defendant is black). She is also biased with

respect to a norm of justice according to which only the guilty should be punished since she punishes innocent black people more often than innocent non-black people. These are just two of the many ways in which one might consider this judge biased on the norm-theoretic account.

The norm-theoretic account allows Kelly to draw some plausible conclusions that appear interesting. For example, Kelly gives a novel explanation of the bias blind spot—our tendency to see bias in other people in ways that we fail to see in ourselves (see also Kelly 2024). The relevant norm, according to Kelly, is a norm of accuracy: people ought to believe things that are true. One is biased with respect to accuracy, call it *truth-biased*, when one's beliefs systematically depart from the truth. Kelly's insight is this: given that we believe as we do, a person who has systematically different beliefs than our own appears to depart from the truth systematically. Hence, in light of our own beliefs, we ought to think this person is truth-biased. At the same time, we ought not to think of ourselves as being truth-biased.

Part 3 of the book is called *Bias and Knowledge*, and it is the most interesting part of the book. It explores epistemological questions related to bias that are to some extent independent of the theory of bias expounded in part 2. Some of its results are: (1) one can genuinely know something even if one is biased with respect to that knowledge; yet (2) a token belief that is a manifestation of bias is (probably) not knowledge; (3) biased people can sometimes be reliable; and (4) it is sometimes rational for laypeople to dismiss the views of experts because they think they may be biased. In what follows I discuss my two main criticisms of the book.

DOES THE CONCEPT OF 'BIAS' HAVE A NATURE?

The word 'bias' is used in many different ways. A wide variety of things can be biased, and they can be biased in multiple senses. Kelly thinks the explanation of this fact is not linguistic. Rather, the diversity of types of bias "reflects something deep about the nature of bias itself" (9). Parts 1 and 2 of the book are devoted to finding out what this nature is—or at least to learning more about the nature of bias, by answering questions such as: What kind of things can be biased? Is there a hierarchical structure between types of bias? I have trouble making sense of this project. Finding out the nature of something presupposes that it has a nature. Does 'bias' have a nature?

First, let me clarify the opposing viewpoint which I find more attractive. The word 'bias' means what we want it to mean. The reason that English has the word is that it has apparent usefulness in the English-speaking world. At any point, we might decide to change its meaning because we find that more useful (or it might evolve naturally). Nothing is lost by such a change, because 'bias' does not refer to something fundamental that exists independently of culture and language, as Kelly suggests. That is not to say that instantiations of bias don't exist, such as biased judges. My position is that the overarching *concept* of bias does not refer to an existing thing, except if what exists is the linguistic reality or the definition chosen by humans.

On this view, nothing is surprising about the fact that so many different things are called biased, in so many ways. That is simply a result of the fact that the word bias means many different things. We would have too many words if each of these meanings was given a separate label. The similarity between the different meanings, combined with the etymological history of the word 'bias', explains that we have a single word for all of these things. Hence, no theory of bias is needed to explain how the different uses of 'bias' relate to each other.

Kelly suggests that my kind of story is implausible:

The fact that we sometimes use the word 'biased' to convey a negative evaluation of the object of our attribution, but in other cases, we don't, seems importantly different from the fact that we sometimes use the word 'bank' to talk about financial institutions and sometimes use the same word to talk about riverbanks. (147)

Here and in other places, Kelly suggests that he thinks a linguistic explanation that the word 'bias' has different meanings is implausible and that we should therefore look for an explanation referencing the nature of bias.

I find some evidence for my position in the fact that not all languages have a similar word. My own native language, Dutch, has no direct translation for 'bias': it is translated in different ways depending on context. In Kelly's view, this should mean that Dutch is missing an important concept, such as we might say when a language has no word for 'cloud', 'truth', or the number '5'. However, I don't think Dutch is missing anything at all. We are doing fine with different words.

Unfortunately, Kelly does not give many reasons for believing that ‘bias’ has a nature. He does give a comparison of his project to the philosophical project of understanding the nature of truth (43–44). About truth, one can ask: which types of things can be true or false? Kelly maintains he is asking such a ‘nature-of’ question about bias: which types of things can be biased? At least in the case of truth, many philosophers would agree that this question can be understood as a nature-of question, rather than a linguistic question (‘To which types of things does the English language attribute truth and falsity?’).

The book is missing an argument that such nature-of questions about bias make sense. Concepts like ‘truth’ and ‘5’ can reasonably be thought to exist independently of human definitions and conventions—and even in those cases, there are philosophers who disagree.¹ However, I don’t know any good reasons to think that the same is true for the concept of ‘bias’.

While this is a major problem with part 2 of the book, the norm-theoretic account of bias is still useful, although perhaps in different ways than the author intended. First, it is useful as a theory for understanding the linguistic practices of attributing bias in English. Second, the concept developed by the theory might be useful in the sense of conceptual engineering. The concept of ‘bias’, as characterised by Kelly, might play a useful role in society, science, and philosophical subjects like social epistemology.

IS THERE A GENUINE EPISTEMIC NORM OF ACCURACY?

In the more epistemological sections of the book, Kelly gives much importance to one specific type of bias that I called truth bias. It plays a central role in his explanation of the bias blind spot (chapter 4), his argument that people should rationally attribute biases to others (chapter 3), his chapter on the epistemology of disagreement (chapter 10), and other chapters. It also features prominently in a recent paper (Kelly 2024). These discussions rely on the view that there is a genuine norm of accuracy, which requires people to have beliefs that are true. I have some reservations that I explain below. First, I am sceptical that there is such a norm. Second, even though truth bias might be a legitimate bias, I think

¹ For an overview of positions that truth does not have a nature (deflationism), see Armour-Garb, Stoljar, and Woodbridge (2023). For an example of a position that mathematical objects have no independent existence (nominalism), see Azzouni (2004).

the book overstates its significance. Other biases and epistemic norms may be more important.

To be clear, it is certainly desirable that our beliefs are true. This widely shared desire explains why we have many of the epistemic norms that we do. However, as a matter of logic, from 'it is desirable that one believes X if and only if X is true' it doesn't follow that 'one ought to believe X if and only if X is true'.

The norm that we ought to believe what is true does not seem to be a genuine epistemic norm to me. This becomes clear if we consider cases in which the truth norm is violated but other epistemic norms are not. Suppose that I put some ice cubes in a jug of water on the kitchen counter. Three hours later, without returning to look at the jug, I form a belief that the ice cubes have melted. (All this time the room temperature stayed at 25 degrees Celsius.) However, it is a well-known fact of statistical mechanics that the ice cubes might still be there, although the chances of this happening are extremely small. According to the above accuracy norm, I ought not to believe that the ice cubes have melted in a possible world in which they have not. However, it seems I ought to believe this in every world in which I have the evidence that I do, including the extremely unlikely worlds in which my belief is false.

The accuracy norm prescribes that the beliefs of people should be dependent on facts that they are unaware of and are in no position to be aware of. Hence, the norm seems to violate an epistemic version of the commonly accepted ethical principle 'ought implies can.' In the ice cube case, I would be required to change my beliefs depending on whether an extremely unlikely event happened that I can't possibly know about or be required to know about. This strikes me as implausible.

Although there seems to be no epistemic norm of accuracy, there is a way out for Kelly. Instead of a norm, there could be a standard of correctness that beliefs should be accurate. After all, biases, according to the norm-theoretic account, involve systematic deviations from genuine norms *or* standards of correctness. Hence, Kelly could maintain that truth biases are biases in this second sense. This seems alright to me. However, while there may be truth biases in this sense, I doubt that they should be given as much significance as Kelly does.

Consider one of his own examples (75-76). Two persons, Left and Right, disagree on politics in a systematic way. Left thinks that Right's positions tend to be too much to the right; Right thinks that Left's positions tend to be too much to the Left. Kelly shows, convincingly, that both

ought to think that the other person is truth-biased, on pain of irrationality. Kelly presents this as an interesting result. But why is this any more interesting than the platitude, ‘I believe I am right, so I ought to believe you are wrong’? A judgement of truth bias could be interesting if it had implications, such as epistemic implications about which substantive propositions one ought to believe. In the situation described here, I don’t think there are any such implications.

Things become admittedly interesting when there is evidence that a person or group of people are truth-biased with respect to some topic. Kelly introduces an example in which scientific studies show that people like Right often have inaccurate opinions. Kelly argues that Left, in such a case, has reason to discount Right’s opinions. This conclusion seems plausible.

In other cases, however, Kelly goes too far with his claims that we can discount others because they are truth-biased. In one example, we imagine that Antonin Scalia’s views on the constitutional legality of restrictions on abortion were insensitive to the truth (suppose we are told this by an infallible oracle). That is, Scalia, who believed abortion restrictions are constitutional, would have also believed that abortion restrictions are constitutional if they were not. Curiously, Kelly claims that Scalia’s opinion in this situation is “evidentially worthless” (216). This strikes me as incorrect.

Kelly’s claim seems to be that Scalia’s view is evidentially worthless only because the view is truth-insensitive. This would imply that Scalia’s truth-insensitive views are evidentially worthless even if Scalia does not violate any other epistemic norms, including norms about responsiveness to evidence. But in that case, the ice cube example is similar. Since I don’t violate other epistemic norms (we assumed), my beliefs, while insensitive to the truth, are properly sensitive to the evidence. For example, if I received additional evidence that my clock malfunctioned and instead only 10 minutes had passed, I would change my beliefs about the ice cubes. On the other hand, my belief that the ice cubes have melted is insensitive to the truth, holding fixed my evidence. However, it is clear that my beliefs have evidential value. My telling you that the ice cubes have melted should increase your own belief that they have.

Turning back to the constitutionality of abortion, suppose similarly that Scalia does not violate epistemic norms besides the (disputed) norm of accuracy. Thus, the infallible oracle’s pronouncement must mean that

Scalia's views would be the same whether or not abortion is constitutional, holding fixed his evidence. However, possible worlds in which abortion is constitutional while Scalia's evidence is the same might be very unlikely, for all we know. In fact, that is what we ought to believe if Scalia is an expert (assuming we have not heard contradictory opinions from other experts). Hence, just as in the ice cube example, Scalia's truth-insensitive view has evidential value.

Another problem for the significance of truth-bias is that we typically don't know that we have truth biases. Rational and justifiable beliefs are often false. Since we are normally not able to know that they are false, there is not much we can do about that. On the other hand, there is a lot we can do about biases related to norms of evidence, such as biases with respect to the norm of having justifiable (rather than true) beliefs. It would have been interesting to see more discussion on that front.

Despite its shortcomings, *Bias* is an impressive work with many interesting discussions on a broad range of topics. While I would not recommend reading it in full, scholars of bias would do well to check what Kelly has to say about this or that topic, particularly in the epistemological part of the book.

REFERENCES

- Armour-Garb, Bradley, Daniel Stoljar, and James Woodbridge. 2023. "Deflationism About Truth." In *The Stanford Encyclopedia of Philosophy, Summer 2023*, ed. by Edward N. Zalta and Uri Nodelman. <https://plato.stanford.edu/archives/sum2023/entries/truth-deflationary/>
- Azzouni, J. 2004. *Deflating Existential Consequence: A Case for Nominalism*. New York, NY: Oxford University Press.
- Kelly, T. 2024. "Bias, norms, introspection, and the bias blind spot." *Philosophy and Phenomenological Research* 108: 81-105. <https://doi.org/10.1111/phpr.12953>

Lennart B. Ackermans is a recent PhD graduate of Erasmus University Rotterdam.

Contact e-mail: <philosophy@ackermans.ch>