

BOOK REVIEWS

Begby, Endre. *Prejudice: A Study in Non-ideal Epistemology*.
Oxford: Oxford University Press, 2021. Pp. 240. \$43.99 (cloth).

You might optimistically hope that by carefully following the dictates of rationality, you'd be guaranteed to avoid having prejudiced beliefs. The main thesis of Endre Begby's *Prejudice: A Study in Non-ideal Epistemology* is that this hope does not withstand scrutiny. More provocatively put, "In sufficiently warped socio-epistemic contexts, there may just be no rational path to true belief: even when our cognitive capacities are operating at their critical best, even when we are doing everything we should (epistemically speaking), there may be no guarantee that the result is not prejudiced" (2). One of the upshots—explored in the final chapter—is that moral responsibility to redress the harms of prejudice cannot depend on having been epistemically culpable. The book is written clearly and approachably, and it nicely compliments other recent work on the epistemic norms for limited agents (esp. Julia Staffel's *Unsettled Thoughts: A Theory of Degrees of Rationality* [Oxford: Oxford University Press, 2019] and Cat Saint-Croix's "Non-ideal Epistemology in a Social World" [PhD diss., University of Michigan, 2018]).

The first portion (chaps. 1–3) clarifies the book's methodology, narrowing its focus in three key ways. First, Begby defines *prejudice* as "a negatively charged stereotype, targeting some group of people, and derivatively, the individuals who comprise this group" (8–9). He focuses exclusively on the epistemic (rather than affective or conative) aspects and resists characterizing prejudiced beliefs as definitionally false. Second, Begby takes *belief* to be the state of having a credence within a sufficiently high range. He takes any functions of outright belief which can't be reduced to credence to be filled by *acceptance*, "an executive decision to treat something as 'actionable'" (23), which does not require thinking it likely to be true. Finally, Begby models *nonideal* epistemic norms as differing from standard idealizing norms only so far as necessary to accommodate both our distinctive cognitive limitations ("endogenous non-ideality") and our limited informational environments ("exogenous non-ideality"). Recast in light of these clarifications, the book's central claim is that if idealizing rational norms are adjusted to (i) permit the reasoning forms made necessary by our cognitive limitations and (ii) count an epistemic state rational when supported by the evidence actually available to the agent, then these norms sometimes permit—or even require—adopting high credence in a negatively charged stereotype targeting a group of people. It's worth noting (if only in passing) that there are two ways to respond to this conditional: One could join Begby in concluding that prejudiced beliefs can be rational. But one could equally well infer that we must more radically revise

Ethics, volume 134, number 1, October 2023.

For permission to reuse, please contact journalpermissions@press.uchicago.edu.

rational norms to accommodate the systematically distortive effects of nonideal contexts; this is the lesson Saint-Croix draws from observations similar to Begby's.

The main arguments for the claim are marshalled in chapters 4–5. Chapter 4 contends that in warped environments the available evidence—especially others' testimony and induction on social categories from limited observations—can rationally favor prejudiced credences. If adults in Johnny's orbit say that women are not as good at math as men, and moreover the top students in his classroom are boys, then "his observations *should* lead him to favor the hypothesis that girls are less [mathematically] adept than boys" (65; emphasis in the original). That Johnny's evidence is misleading is unfortunate, but this does not make it less rational to proportion his credences to it. Chapter 5 argues that it is possible to rationally retain a prejudiced belief despite counterevidence: first, because prejudiced beliefs are often encoded as generics, not falsified by (even substantial numbers of) counterexamples; and second, because even if conditionalizing appropriately on counterevidence lowers one's credence in the prejudiced hypothesis, it may still fall within the relevant range for belief. Chapters 6–7 highlight ways in which the social dynamics of testimony and inferences about apparently "socially normal beliefs" can make it especially difficult to rationally unseat prejudiced beliefs once acquired.

The final three chapters take a broader look at the relationship between epistemic and moral norms. Chapter 8 discusses algorithmic risk assessment tools as a case study illustrating how rampant background injustice can result in conflict between what epistemic norms prescribe and what would be fair or socially just. Chapter 9 argues that neither appealing to moral encroachment nor doxastic wronging can bring rational and ethical norms into harmony; it remains possible for evidence to strongly support beliefs that are deeply morally harmful. The final chapter argues that rather than insulating us from responsibility for our prejudiced beliefs (and for the harms they cause), the lesson is that moral responsibility cannot depend on having been in a position to know better. This is a quite important claim in the context of the book. Defenses of the rationality of prejudiced beliefs are often leveraged to deflect moral responsibility or criticism, but Begby's project is the opposite: to illustrate that very little of ethical significance follows from the fact that someone conformed to the norms of rationality. Ethical norms are independent of epistemic norms; they can come into conflict, and abiding by one set is no guarantee against violating the other. The book touches on many themes in feminist and nonideal political philosophy (and explicitly takes up some issues centrally discussed by Miranda Fricker, Michele Moody-Adams, Charles Mills, and José Medina) but engages primarily with recent literature in informal epistemology. It would doubtless provide rich material for a graduate seminar—especially together with Staffel's discussion of how Bayesian constraints on rational credence revision and conceptions of the value of accuracy need to be reformulated for imperfect agents.

One of the book's signature virtues is its focus on tracing the contours of the rational norms appropriate given our limitations—norms which identify the sense in which an agent who proportions her belief to the available evidence is getting something right, even if it turns out that her evidence was misleading. It is laudably committed (particularly in chaps. 2–3) to outlining the cognitive constraints of beings with brains and communities like ours, drawing on recent

work in psychology and social epistemology to argue that because social learning and categorization are necessary for timely completion of “a variety of obligatory cognitive tasks” (33), the rational norms applicable to beings like us cannot simply demand that we refrain from stereotyping or social induction, full stop. But the level of un-idealization in the project is uneven, and I found myself wondering whether accommodating our cognitive limitations demands a more thoroughgoing departure from standard idealized epistemology.

Begby sometimes pitches his thesis as the claim that abiding by rational norms is consistent with forming prejudiced beliefs in some contexts, and sometimes as the stronger claim that in such environments one is rationally required to form them. The argument for the weaker claim is clear and relatively straightforward given the book’s definitions of the key terms. When forces ranging from chance to injustice distort the evidence, merely proportioning one’s credences to the available evidence is no guarantee against becoming confident in negative stereotypes targeting social groups. But the argument for the stronger claim depends on assuming that we are rationally required to treat as relevant “any bit of information that increases or decreases the probability of a hypothesis under consideration” (17). This assumption is what underwrites the claim that Johnny is rationally required to treat the fact that most successful math students in his class are boys as evidence—however weak—in favor of believing that girls are less adept at math. If rational norms require us to respond to skewed distributions in small samples (or testimony from others in an insular community) as evidence, it is clear enough how following these norms in sufficiently warped socio-epistemic contexts could lead us to form high credence in prejudicial stereotypes.

It is standard in idealized epistemology to accept both this capacious interpretation of evidence and a norm requiring agents to update on all evidence they encounter. And it makes sense to require unlimited agents to attend to all information, even when its justificatory force is marginal, or its significance ambiguous. In a sufficiently large body of evidence, misleading information will eventually be outweighed and spurious correlations washed out. But it is a hazardous rule to give to agents like us—in part due to the very same cognitive limitations Begby invokes to motivate his project. For one thing, our attention comes at steep opportunity cost; we must prioritize when allocating our cognitive resources. For another, when dealing with ambiguous or weak evidence, we are likely to overestimate its import or misidentify what it supports. And when working with only a small fraction of total information, it can be extremely hard to recover from these sorts of mistakes. Agents with cognitive limitations like ours will often do better—that is, their credences will be more coherent and accurate—if they ignore information when they aren’t in a position to determine exactly how it bears on their inquiry. So I think there is room to doubt that rational norms for nonideal agents will forbid this.

To motivate the claim that it is at least rationally permissible for epistemically limited agents to sometimes discard probabilistically relevant information, imagine you’re attending an Object Tossing Expo where four kinds of objects (coins, buttered bread, beanbags, and horseshoes) will be tossed. Every object has a fancy side and a plain side, each toss is of a distinct object, and every object is painted one of three colors (red, blue, green). You’ve seen fifty tosses so far: twenty landed

fancy side up, twenty-eight landed plain side up, and two were edge cases. The next object to be tossed is a green coin, and you entertain two hypotheses:

H: This coin will land fancy side up.

~H: This coin will land plain side up.

The world is buzzing with details, and you've observed some interesting patterns in the tosses. An astonishing eight of the nine green objects tossed landed fancy side down—including every single piece of buttered bread. Just ten of the tosses so far have been coins; seven landed fancy side up. With all this information at your disposal, which hypothesis (H or ~H) should you favor? All of the observations have some bearing on H's probability. But are you rationally required to use them all to set your credence concerning whether the coin about to be tossed will land fancy side up?

If we say no, we're invoking a narrower conception of evidence than Begby assumes. But if we affirm that it's all relevant evidence, you face the reference class problem. To get around it, you need to know which object groupings are more predictive with respect to H; this introduces space for a rational permission to not treat as evidence observations in any category for which one lacks evidence of predictive power. Either way, there's space to deny the claim that someone who ignored this sort of information "would essentially be guilty of throwing away evidence for no good reason" (65). Rather, you can rationally ignore all the bread, beanbags, and horseshoes, letting your credence in H be informed exclusively by the distribution of past coin tosses—even the ones that weren't green. In fact, you could even rationally ignore the previous coin tosses. Begby acknowledges this for relevantly similar cases but maintains that it is only because you have special reason to doubt its relevance: "coin tosses are designed to have a random outcome, and that's how we know that we should resist our inductive impulses even in a case where a first series of observations skews in favor of heads. What would be our grounds for thinking that similar restrictions apply in social cognition?" (67). For limited agents, though, the opposite default seems rational: we should refrain from induction on a category unless we have special reason to think that it is relevant. This means (contra Begby) that Johnny is not rationally required to respond to his transparently tiny sample by increasing his credence in the generalization that girls are less mathematically adept than boys. He in fact has very good reason to dismiss it—similar to your reasons for dismissing the green object tosses from relevance when setting your credence in H. Absent positive reason to think that a category has predictive power, treating facts about other members of the category as evidence concerning unobserved instances is likely to lead to errors from which you cannot easily recover. Better to take yourself to have no evidence at all. In brief: if we are rationally permitted to doubt the evidential relevance of past coin tosses for predicting the outcomes of future coin tosses, plausibly a similar skepticism is at least permitted to Johnny. If so, rationality does not require that even someone in circumstances as unfortunate as his form prejudiced beliefs.

This does not threaten the weaker reading of the book's thesis. Induction on a small set of social observations is not the only route to prejudiced belief, and it's possible to have a reason to think that a social category is inductively relevant. But

it does prompt a question: if we are persuaded that the rational norms applicable to agents like us must accommodate our cognitive and situational limitations, what other standard assumptions should we revisit before we can draw conclusions about the normative demands of nonideal epistemology?

The decision to treat outright belief as “really just vague talk for having a credence somewhere in a certain range” (22) raises a similar issue. It might be that in idealized agents with unlimited computational power belief serves no role that blocks reducing it to credence—though this is controversial. But agents with our constraints can’t sustain open inquiry into every proposition we’ve ever entertained. And while we can use our credence in p to make a calculated bet hinging on p , we find it at best cognitively unwieldy—often impossible—to use iterated conditional probabilities when navigating the world. That is, when our confidence in p is a function of our credence in q and r , while r depends on s , and so on, timely completion of the cognitive task demands that we simplify rather than perform the complex calculation. So, among the propositions we think likely to be true, we treat some as fixed points to structure our reasoning and inference and to define decision problems for our practical deliberation. This is a binary attitude thoroughly compatible with being less than certain but exclusively taken toward propositions one thinks are most supported by the evidence, as well as sufficiently justified to structure reasoning, inference, and action in the context—acceptance as Begby characterizes it does not fit the bill.

As a consequence, the otherwise rich discussion of moral encroachment in chapter 9 talks past at least some intended interlocutors. Many of its advocates pitch moral encroachment as affecting whether simplifying belief is rational, not whether your credence in p should reflect your evidence, nor (merely) whether you should act on p . Since the book focuses exclusively on credences and doesn’t really engage with whether nonideal epistemology must posit something to fill the simplifying role of belief, it’s hard to evaluate its charge that “what [moral encroachment] offers is not a moral constraint on belief formation, but a moral constraint on action (broadly construed, including the executive decision to treat certain beliefs as actionable)” (162). I think that Begby is right that embracing moral encroachment does not resolve the questions that are his focus. Encroachment offers no guarantee against following misleading evidence to high credence in negatively valenced stereotypes. But encroachment does imply that one cannot rationally adopt the simplifying attitude belief on the basis of those credences when doing so poses a high risk of harming or wronging others.

The book does an excellent job motivating inquiry into what a system of epistemic norms for agents with our specific cognitive and situational limitations would look like, and it forcefully contends that we cannot condemn an inference solely for involving stereotyping. But the way it engages with this question leaves a lacuna in the argument that invites further investigation. To draw conclusions about what rationality requires from nonideal agents, we must first give similar attention to the rest of the framework: asking how our limitations affect how we ought to respond to information, as well as our cognitive ecology more generally.

RENÉE JORGENSEN

University of Michigan at Ann Arbor