Michael Brownstein and Jennifer Saul, eds., *Implicit Bias and Philosophy*, Volume 1: Metaphysics and Epistemology. Oxford University Press 2016. 336 pp. $65.00 USD (Hardcover ISBN 978019871324).

Implicit Bias and Philosophy Volume 1: Metaphysics and Epistemology, edited by Brownstein and Saul, is the result of several conferences held at the University of Sheffield in 2011 and 2012. This edited volume brings together established and emerging scholars for novel contributions to this topic. These articles present challenges for those of us thinking about the nature of racism and sexism, but also demonstrate how thinking about racism and sexism impacts philosophy more generally. The contributions in this volume are insightful, (for the most part) well argued, and accessible.

Brownstein and Saul's introductory chapter (part of which is adapted from Brownstein's *Stanford Encyclopedia of Philosophy* article 'Implicit Attitudes') nicely introduces readers to the paradigmatic empirical studies essential to understanding the papers in *Implicit Bias and Philosophy Volume 1: Metaphysics and Epistemology*, making this volume accessible for students. Another strength of this volume, likely due to the multiple conferences and workshops by the Implicit Bias Project, is the way in which contributors interact with each other's papers, especially when disagreement arises, giving the book a dialogue feel.

Implicit Bias and Philosophy is divided into two parts, the first on the cognitive architecture of implicit bias, the second on its epistemology and epistemic consequences. Somewhat surprisingly, only two authors utilize a dual-process approach to implicit bias. Frankish's article is an application of his own version of dual-process theory to implicit bias. He offers some predictions for his theory, but it would have been nice for interaction with the numerous criticisms of dual-process theory, especially since few of the other authors in this volume seem to adopt dual-process theory. Mallon situates implicit bias within the personal/subpersonal distinction and utilizes a dual-process framework, but he rightly points out that the personal/subpersonal distinction does not neatly map onto the type-1/type-2 distinction. He argues that stereotype threat (but not necessarily implicit bias generally) occurs at the personal level, rather than the subpersonal, as often assumed.

Huebner, Holroyd and Sweetman, and Machery's articles offered novel ways of understanding implicit bias, with the latter two providing reason to be skeptical that 'implicit bias' is a phenomenon at all. Huebner offers a novel cognitive architecture according to which implicit biases arise from an interaction between three kinds of systems: Pavlovian (associations between stimulus and response), model-free (associations between action and outcome), and model based (counterfactually based). An important upshot of his account is that we cannot remove implicit bias until 'we eliminate the conditions under which they arise' (71). That is, until our society *is egalitarian,* we will be unable to truly think *as egalitarians.* Machery argues against the wide consensus that implicit biases are attitudes on multiple grounds, including that there are low correlations across the various measures for bias and that 'while indirect measures do predict behavior, they are poor predictors' (119). Machery goes so far to argue that 'there is no such thing as implicit racism or implicit sexism' (115): those claiming to be egalitarian but succumbing to implicit bias simply are not egalitarian. Machery offers a trait (dispositional) picture of attitudes to replace the old theory. Holroyd and Sweetman police the 'implicit bias' concept by pointing to two concerns for its (overly) general use. 'Implicit bias' is not a kind, and this matters because how we intervene on biases depends on their function and structure. Holroyd and Sweetman argue that implicit biases do not all function the same way, and how we can combat them depends on the function of implicit bias. Holroyd and Sweetman and Machery's articles are radically revisionary, but well argued. If you only want to read a couple of

articles from this volume, read theirs. A nice feature of each article in Part 1 is that each author suggests practical implications for combating pernicious bias based on their account (or skepticism) about implicit bias.

There is less interaction among the authors of the epistemic portion of the book. This may be due, in part, to the more divergent topics in this section. Antony and Madva both attempt to solve epistemic problems resulting from implicit bias. Antony takes issue with Saulish skepticism: given that we are biased, how can we trust our epistemic agency? Antony helpfully situates the problem within a naturalized epistemology: because we do not have infinite time and processing power, 'we should not strive to put aside all bias,' even from an epistemic perspective (161). Unfortunately, because of the way society is structured, many of the pernicious biases turn out to be useful (185). Antony's solution to Saulish skepticism make the moral-epistemic dilemma all the more pressing. Madva attempts to reply to Gendler and Egan's tragic dilemma by saying that agents might access stereotype information only in those cases in which they are epistmically relevant. One might wonder if this is empirically possible, since access to stereotype information is automatic. Furthermore, Antony's version of the dilemma presents further challenges for Madva, since she argues that most pernicious biases do track contingent truths.

Goguen and Hundleby's articles offer novel ways of understanding implicit bias. Goguen argues that the problems of stereotyping go beyond what has previously been called 'stereotype threat.' It is unclear whether Groguen has really demonstrated that we need "a broader account of stereotype threat" (216) or whether she has explicated additional threats of stereotyping. It might be advantageous to use 'stereotype threat' in the restricted sense and simply add terms for phenomena closely related to it, such as avoiding those tasks one's group is stereotypically bad at. Hundleby situates implicit biases within argumentation theory as instances of status quo bias. It is an interesting idea, but the central claim is under argued.

The final two articles critically examine or conduct studies to determine why women are less represented in the STEM disciplines and philosophy. While both are timely, they feel like outliers in this volume. Lee takes issue with Ceci and Wendy's (2011) claim that "gender discrimination in journal review, grant funding, and hiring [in the STEM disciplines] are 'no longer valid'" (265). Bella, Miles, and Saul test a number of hypotheses concerning why women are underrepresented in philosophy. The title 'Philosophers Explicitly Associate Philosophy with Maleness' is somewhat misleading. Although this was a finding, it could just as easily have been titled 'Philosophers do not Implicitly Associate Philosophy with Maleness' or "Increasing Readings from Female Authors does not Alter 'Male'/'Female' Associations with 'Philosopher'." The data from Bella, Miles, and Saul raises more questions than it answers, and opens "several promising lines for inquiry of further" empirical investigation (304). I hope we will see more data from them soon.

Implicit Bias and Philosophy Volume 1: Metaphysics and Epistemology is a must read for anyone working on the philosophy of race and gender. Philosophers purely interested in cognitive architecture would do well to read Part 1, as biases have much to teach us about the structure of the mind. Epistemologists would do well to read the first two articles in Part 2. Every philosopher should read Bella, Miles, and Saul's article and think about implicit bias's effect within our own discipline. After all, the hope is that studying implicit bias can help us overcome the pernicious implicit biases.

Joshua Mugg, Indiana University