

Michael Brownstein and Jennifer Saul (editors)
Implicit Bias and Philosophy, volume 2: Moral Responsibility, Structural Injustice, and Ethics
Oxford: Oxford University Press, 2016 (ISBN: 978-0-19-876617-9)

Reviewed by Katherine Puddifoot
University of Birmingham

Katherine Puddifoot works at the intersection of philosophy of psychology, epistemology, and feminism. Her research focuses on epistemological issues arising from research on cognitive biases, including implicit bias and stereotyping, and memory distortion. She is currently research fellow on the ERC-funded Project PERFECT at the University of Birmingham. She has previously occupied philosophy posts at the University of Glasgow and the University of Bristol.

**

Implicit Bias and Philosophy, volume 2: Moral Responsibility, Structural Injustice, and Ethics is the second in a two-part collection addressing philosophical issues arising out of psychological research on implicit bias. Implicit biases are automatic and unintentional forms of cognition that associate members of social groups with traits or affective responses, and sometimes operate under the radar of consciousness.

The collection derives from a series of workshops held at the University of Sheffield in 2011-2012. The workshops played a significant role in shaping the direction of research in the philosophy of implicit bias. Because of the length of time between the workshops and the publication of the volume, many of the chapters in the collection are already influential in the field, and the collection as a whole comprehensively covers a wide range of important philosophical and practical issues.

The second volume covers a wide range of moral and ethical issues relating to implicit bias. The first part focuses on whether people are morally responsible for implicit bias and, if so, how it is possible to account for this responsibility. The second part addresses the relationship between structural injustices and attitudes of individuals. The third part includes chapters on a diverse set of issues relating to implicit bias, ethical theory, and practical measures to tackle bias.

Section 1 includes a series of discussions of how people can be morally responsible for implicit biases if they are unaware of, cannot control, or do not endorse the operation of the biases. Natalia Washington and Daniel Kelly (chapter 1.1) argue that people can be morally responsible for implicit biases even if they are unaware of the influence of biases on their behavior and cannot control them at the time of their influence. If there is knowledge available in a person's epistemic environment about implicit bias, and they occupy a social role (for example, hiring-panel member) in which implicit bias is clearly relevant, then they can be held responsible. Washington and Kelly conclude that their argument justifies a form of externalism because moral responsibility is determined by something "outside the head," that is, the availability in one's environment of knowledge gained through empirical research.

Joshua Glasgow (chapter 1.2) develops an account of moral responsibility that is consistent with people being morally responsible for implicit biases from which they are alienated as a result of disavowing them. The account, *context sensitive variantism*, says that whether a person is blameworthy for an action or attitude from which they are alienated depends on its context: whether it is harmful. Sometimes alienation leads to exculpation from responsibility (for example, in the case of an unwilling addict or petty thief), but sometimes it does not (for example, in the case of a cheating partner or implicit bias). It is when an action or attitude violates a nonnegotiable value (for example, those against infidelity or implicit bias), which impairs an agent's relationships, that the agent is not exculpated from responsibility.

Robin Zheng (chapter 1.3) argues that it is important to distinguish between two types of moral responsibility when discussing whether people are morally responsible for implicit biases. According to the *attributability account* of responsibility, we are not responsible for behaviors arising from implicit biases if we do not endorse the biases and have done all that we can do to prevent them from influencing us. This is because the biases are not attributable to our agency, volition, or character. However, we can hold people morally responsible in such cases because they meet the conditions for *accountability responsibility*; they can be expected to make amends for the negative consequences of their biased behaviors. For Zheng, applying the notion of accountability responsibility to cases of implicit bias facilitates an appropriate level of moral criticism. It fosters the types of relations needed within a moral community by holding people accountable when they behave poorly, even if their thoughts and actions are not attributable to them.

Maureen Sie and Nicole van Voorst Vader-Bours (chapter 1.4) argue that when stereotypes operate without the agent's awareness of their operation or of the harm that is done, responsibility for implicit bias comes from being members of the society that creates and sustains the stereotypes. They argue that stereotypes are collectively maintained and can be eradicated only if people collectively reject the stereotypes. As stereotypes can be eradicated only by the collective, the focus in discussions of moral responsibility should be on the responsibility that people have as a part of a collective.

Luc Faucher (chapter 1.5) proposes revisions to three ideas relating to moral responsibility. First, the notion that people are morally responsible only if they are aware of the operation of their bias should be revised to reflect the views of people who have been subject to discrimination, who ascribe responsibility to those who discriminate even when the discriminators are not aware of being biased. Second, the idea that people are responsible for their actions only if they stem from *consciously chosen* choices or ends should be revised because empirical findings suggest that people's attitudes and motivations can be reflected in their implicit responses as well as their consciously chosen attitudes. Third, the idea that control requires awareness should be revised because empirical findings show that it is possible to adopt long-range strategies that ensure that if we encounter a situation in which we would be likely to be biased, the biases are less likely to manifest. These long-range strategies can control the influence of bias even when people are unaware that they might be biased in a particular situation.

The chapters in section 1 on moral responsibility provide interesting new ways of conceptualizing both responsibility for implicit bias and the contours of debates about moral responsibility.

Glasgow, Zheng, and Faucher provide innovative, revisionary accounts of moral responsibility, highlighting shortcomings of the "folk" conception of moral responsibility, and challenging philosophical assumptions that underlie some philosophical work on moral responsibility (including for implicit bias). Questions remain about how far accounts of moral responsibility should be driven by folk conceptions, and to what extent folk conceptions should be revised in light of empirical findings. However, the discussions from these authors provide interesting suggestions about how the folk conception could be retained but modified, even while it is acknowledged that there are inconsistencies in the judgments made by the folk.

Meanwhile, the chapters from Washington and Kelly and by Sie and van Voorst Vader-Bours effectively shift attention away from focusing solely on the knowledge that individuals have about their own implicit biases toward the knowledge that they *should* have and the responsibility that they have *as members of a society* in which stereotypes are in operation. Some psychologists and philosophers have raised doubts about the appropriateness of defining implicit bias as unconscious biases, so the question raised by Washington and Kelly about whether we can be responsible for implicit biases if we lack awareness seems less pressing than it would be if implicit biases were consistently defined in terms of a lack of awareness. Nonetheless, Washington and Kelly's discussion is valuable as long as there are some implicit biases that operate under the radar of consciousness (which there is currently little reason to doubt). With respect to Sie and van Voorst Vader-Bours's account, I worry that where stereotypes *do* operate without subjects' awareness, there will be little value in ascribing responsibility to the subjects, either as individuals or as a part of a collective, until they have undergone education on implicit bias. Only once people have been educated about implicit bias are they likely to engage in the collective action required to reduce stereotyping. But then it seems they will be responsible not primarily because they are a part of the collective responsible for the stereotyping, but because they are aware of the phenomenon.

At the beginning of section 2, Lawrence Blum (chapter 2.1) critiques the stereotype-threat literature, which has emphasized how students can underperform when a stereotype associating their social group with poor performance on a task is salient to them. For Blum, it is important to distinguish useful generalizations about groups, for example, about the inequalities that they face, from stereotypes, but the stereotype literature does not do this. Blum argues that the stereotype-threat literature also underplays the role of social structures in sustaining stereotypes that lead to disadvantage. Meanwhile, awareness of the injustices faced by one's social group could provide a good means of combating one's tendency to doubt one's own abilities due to stereotypes about one's group. The stereotype-threat literature therefore depoliticizes a political issue and distracts attention from promising ways to tackle educational disparities.

Anne Jacobson (chapter 2.2) also addresses the need for structural changes to combat the negative effects of bias and discrimination. Jacobson presents examples that show that changes to attitudes are often ineffective if social structures, such as white

networks, are not changed as well. However, Jacobson rejects calls to shift attention away from implicit biases altogether and to engage instead in group protest and agitation in order to change institutions. While acknowledging the need to change institutions, Jacobson emphasizes that we need not neglect attitudes or we risk reinventing unjust institutions.

Blum's contribution to the volume is only loosely linked to the main topic of the volume, implicit bias, and its attack on the stereotype-threat literature appears in places to be unjustified, criticizing the research program for doing important work because it does not do other important work. Nevertheless, Blum's chapter provides contextualization of the stereotype-threat literature, encouraging people to think beyond this widely discussed phenomenon when looking for sources of educational disparities. The emphasis placed on the need to balance changes to institutions with changes to attitudes found in the work of both Blum and Jacobson is of great importance.

Section 3 begins with Clea F. Rees (chapter 3.1) presenting a defense of virtue ethics, according to which people are encouraged to develop virtuous habits. Implicit bias research seems to challenge virtue ethics because it suggests that vicious patterns of thought are automatically habituated in a way that is not influenced by the deliberation of the agent. However, Rees discusses psychological findings suggesting that processes of deliberation can lead to the automatization of egalitarian goals and commitments, resulting in people displaying low levels of implicit bias. Psychological findings are therefore taken to support rather than undermine virtue ethics by showing that a person can act in a virtuous way and display low levels of implicit bias through habit.

Michael Brownstein's chapter (chapter 3.2) highlights the importance of context in determining whether a person displays implicit bias. A number of psychological findings are discussed: for example, findings suggesting that de-biasing strategies are often effective only in the specific context in which they are initially adopted, and findings showing the impact of mood on the manifestation of implicit bias. In light of the findings, it is proposed that people aiming to tackle implicit bias should focus on the features of their ambient environment and how these features shape their attitudes, putting themselves in the correct relationship to their environment.

Samantha Brennan (chapter 3.3) discusses micro-inequities: small and often subtle inequalities that can occur without the intention or detection of the perpetrator, for example, dismissive facial expressions or condescending tones of voice. Implicit biases often overlap with micro-inequities where the biases lead to micro-inequities. Brennan identifies the wrongness of micro-inequity in the harm that results from it. As a result, whether a micro-inequity is wrong is determined by the context: sometimes a micro-inequity will cause no harm, but on other occasions, for example, after a person has encountered a multitude of micro-inequities, there might be harm, such as the person suffering from depression. For Brennan, as the context determines the wrongness of the micro-inequity, it is proper to focus not on individual acts but rather on the context when looking to reduce or eliminate the harm. Changes to the contexts in which micro-inequities might occur, such as the environment in educational institutions, are most likely to reduce or eliminate the harm, and therefore the wrong, of micro-inequities.

In the final contribution to the volume, Katya Hosking and Roseanne Russell (chapter 3.4) highlight limitations of the British legal system that mean that it does not cope adequately with implicit bias. The authors argue that the legal system places the burden of proof on claimants that it will often not be possible to meet in cases of implicit bias. Moreover, there is a general assumption that differences in people's situations come about as a result of free choice, and therefore a lack of acknowledgment of the social changes that are needed for true equality, including tackling implicit bias. Some equality laws promoting consideration of equality in decision-making in the public sector have the potential to tackle implicit bias. However, the laws are not well enforced, are restricted to public and not private bodies, and stipulate that concern should be taken only for equality and not that equal outcomes are required. The authors provide some positive proposals about how the legal system can be changed so that it is better able to handle cases of implicit bias.

The chapters in section 3 of the volume are more varied than those in the other parts of the book. However, each presents important insights, and there are some significant connections between them. Rees provides an insightful look at psychological models of bias, goal-formation, and motivation that are often neglected by philosophers convinced by the dual-processing explanation of automatic cognition. Meanwhile, by emphasizing the role of context in determining whether behaviors manifest implicit bias, Brownstein succeeds in showing that implicit bias does not inevitably follow from the perception of a person's group membership. It should be noted, however, that an ethics of implicit bias that focuses on agents putting themselves in the correct relationship to their ambient environment is unlikely to satisfy those who emphasize the importance of changing one's implicit attitudes. Brennan's work is connected to Brownstein's due to their shared emphasis on the need to attend to the context in which biased behaviors manifest. It is a merit of Brennan's approach that she emphasizes how micro-inequities become harmful due to their cumulative effects, but the reader who thinks that people are accountable for even small harms that they cause might be left unsatisfied that no explanation is given of how this can be.

Overall, *Implicit Bias and Philosophy, volume 2: Moral Responsibility, Structural Injustice, and Ethics* is comprehensive, insightful, and often innovative. It is required reading for all philosophers working on implicit bias, and all psychologists interested in the theoretical implications of psychological research on implicit bias. Philosophers working on moral responsibility might be interested in the diversity of approaches to responsibility stimulated by implicit bias research. As the issues discussed in the book have important practical implications, they should be of interest to a non-academic audience, including policymakers.