Introduction

Tobias H. Stark and Jon A. Krosnick

To begin a tour of research on implicit bias, the construct must be defined conceptually and operationally, and Section 1 does just that. As we shall see, the accumulated literature has been characterized by definitional divergences that merit investigation and resolution.

First, it is worth noting that the notion of "implicit" in implicit bias has been conceptualized in at least four ways: (1) that people who harbor implicit bias do not know that they are biased, (2) that implicitly biased people are aware of their bias but are unwilling to admit to that bias publicly, (3) that people are aware of their bias and suppress its impact on their behavior, and (4) that the method of assessing implicit bias hides from participants what is being assessed.

The MODE model offers a valuable conceptual framework for understanding implicit bias (Fazio, 1990; Olson & Fazio, 2008). According to this model, attitudes are associations in memory between an attitude object and evaluations. Implicit bias is an automatic association or an automatically activated evaluation that will affect thinking and action when a person is unable and unmotivated to suppress its impact. But when people have sufficient motivation and opportunity, they may choose to control themselves in ways that minimize the impact of automatic activation. This "dual-process" idea, introduced here, appears in many chapters of this handbook. It has clear links to the concept of aversive racism (Chapter 4), anti-implicit-bias training programs (see Section 5), and cultural interpretations of implicit bias (see Section 6).

The MODE model offers an explanation for modest correlations between explicit racial prejudice and implicit racial prejudice: motivation and opportunity to control prejudice. When people are not motivated to control their prejudice or do not have sufficient resources to do so, a strong correlation between explicit and implicit prejudice is expected. Thus, the modest correlations often observed are attributable not to the inadequacy of the notion of implicit prejudice but rather to the conditions necessary in order to observe its consequences.

The concept of aversive racism offers another useful perspective (Dovidio et al., 2017; Gaertner & Dovidio, 1986). The starting point of this notion is that well-intentioned people genuinely endorse egalitarian values and believe that they are not prejudiced against Black people, but at the same time possess conflicting, nonconscious, negative feelings toward those individuals. Discrimination against African Americans is, according to this perspective, often the consequence of unintentional behavior. People high in aversive racism are characterized by low explicit prejudice but high implicit prejudice. Due to the lack of awareness of their own bias, aversive racists are especially likely to manifest discrimination.

According to the aversive racism perspective, people who think of themselves as being without bias might nonetheless engage in subtle forms of discrimination if they can

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 11:02:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108885492.005

justify such behavior based on a non-race-related socially acceptable standard. This links aversive racism closely to ideas in the New Racism tradition such as Modern Racism and Symbolic Racism that re-emerge in Section 6 of this handbook.

As we shall see, some voices have challenged these conceptualizations of implicit bias. For example, some central claims made early on about implicit bias have not been substantiated by empirical evidence (e.g., Mitchell & Tetlock, 2017). And the conceptualization of implicit bias has changed over time, leading to some confusion (Blanton et al., 2015). And the notion of implicit bias making its way into popular discussions of prejudice outside of academia has often lost some important nuances. For instance, when members of the public have received feedback on the meanings of their implicit bias scores, this feedback has not always been precisely in line with the accumulated scientific evidence.

Concern has also been expressed that public discussions of implicit bias have sometimes gotten out ahead of the development of solid empirical foundations for all claims made (Mitchell & Tetlock, 2017). For instance, claims about the powerful, unavoidable, and universal impact of implicit prejudice on discriminatory behavior have not always been supported by the accumulated body of empirical evidence. Critical voices have therefore expressed concern that the public's understanding of implicit bias may not be in line with the science, a notion explored directly in Section 7.

In sum, this section provides an overview of conceptualizations of implicit bias and offers a summary of central findings in the literature. Given the prominence of Project Implicit in this literature, special attention is being paid to the history of this project and its central findings. Other operationalizations and conceptualizations of implicit bias are also spotlighted, as are some critical views, setting the stage for the next sections of this book.

References

Blanton, H., Jaccard, J., Strauts, E., et al. (2015). Toward a meaningful metric of implicit prejudice. *Journal of Applied Psychology*, *100*(5), 1468–1481. https://doi.org/10.1037/a0038379

Dovidio, J. F., Gaertner, S. L., & Pearson, A. R. (2017). Aversive racism and contemporary bias. In C. G. Sibley & F. K. Barlow (Eds.), *The Cambridge Handbook of the Psychology of Prejudice* (pp. 267–294). Cambridge University Press.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (pp. 75–109). Academic Press.

Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, Discrimination, and Racism* (pp. 61–89). Academic Press.

Mitchell, G., & Tetlock, P. E. (2017). Popularity as a poor proxy for utility. In *Psychological Science Under Scrutiny: Recent Challenges and Proposed Solutions* (pp. 164–195). Wiley. https://doi.org/10.1002/9781119095910.ch10

Olson, M. A., & Fazio, R. H. (2008). Implicit and explicit measures of attitudes: The perspective of the MODE model. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the New Implicit Measures* (pp. 19–63). Psychology Press.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 11:02:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108885492.005