Thinking Structurally About Implicit Bias: Some Peril, Lots of Promise

Natalie M. Daumeyer, Julian M. Rucker & Jennifer A. Richeson

Routledge
Taylor & Francis Group

Check for updates

Thinking Structurally About Implicit Bias: Some Peril, Lots of Promise

Natalie M. Daumeyer, Julian M. Rucker, and Jennifer A. Richeson

Department of Psychology, Yale University, New Haven, Connecticut

In the wake of the U.S. civil rights and women's rights movements, expressing some forms of explicit bias became considerably less acceptable, and many, if not most, Americans openly embraced egalitarian values regarding different social categories. Although many explicitly discriminatory laws, especially based on race, religion, gender, and national origin, were struck down and/or repealed, other forms of discrimination based on these group memberships persisted. In part to understand this discrepancy, researchers began to focus on the role of *implicit bias*— automatic and, perhaps unconscious, processes of mind that promote biased judgments, decisions, and behavior. In other words, despite good intentions and/or egalitarian values, implicit beliefs and/or associations often result in biased thinking, information processing, and actions nonetheless.

As Payne, Vuletich, and Lundberg (this issue) note, the study of the role of automatic processes in the manifestation of discrimination and maintenance of disparities in contemporary society has been quite fruitful and a necessary corrective to the idea that discrimination is typically, if not necessarily, born of individual intent and, further, a matter of individual-level animus toward specific societal groups. One unintended consequence of this focus, however, has been what can only be described as an overwhelming focus on the implicit biases of individual actors in current discourse regarding contemporary prejudice. That is, the field's focus on implicit bias largely shifted the attention from the role of individual hearts (i.e., feelings) in the production of racial and other forms of bias to the role of individual minds in its production. The role of societal and structural factors in stereotyping, prejudice, and discrimination, in other words, continued to be largely ignored.

In their bias of crowds model, Payne and colleagues (this issue) argue that this overwhelming focus on implicit bias as a stable individual difference variable has been misleading if not counterproductive. Although such an orientation made sense, especially given the effort to contrast implicit bias with explicit bias, it may have nevertheless resulted in what Payne and colleagues refer to as "puzzles" in our understanding of implicit bias. Instead, the bias of crowds model forwards the idea that implicit bias(es) are largely the product of situational factors that make certain associations more accessible than others, yielding a pattern of instability and low predictive utility at the individual level but relative stability and striking predictive utility within situations and contexts and, under certain conditions, across time at the aggregate level (i.e., at the level of states). The model offers several novel predictions and

implications. Here, we offer two additional potential implications that we believe are relevant to this approach and, considered in tandem with the Payne et al. model, suggest the need to rethink our ideas regarding the roles of implicit and explicit bias in the production, acceptance, and remediation of discrimination in contemporary American society.

Reasoning About Implicit versus Explicit Bias

The bias of crowds model (Payne et al., this issue) argues that the dominant conceptualization of implicit bias needs to shift. Implicit bias–thought of as mental associations that individuals learn early on; that are resistant to change; and perhaps that operate unconsciously and, thus, outside of awareness–should be perceived as mental associations that are triggered in specific situations and "pass through" individual minds. From this perspective, implicit biases are relatively malleable so long as individuals find themselves in situations and contexts that differ from one another in ways that are relevant to specific biases. Although this shift will undoubtedly require and inspire new directions for research on implicit bias, it is likely also to reshape the ways in which people reason about implicit bias(es) as a cause of discriminatory outcomes.

Unlike the overt expression of some forms of bias (e.g., racism, sexism, xenophobia) that is generally denounced (Crandall, Eshleman, & O'Brien, 2002), we know very little about how people respond to evidence of the influence or expression of implicit bias, be it by themselves or by others. How people respond to the role of implicit bias in either their own or others' judgments and behaviors, however, is certain to depend, at least in part, on how implicit bias is conceptualized. Consequently, it is important to consider the potential implications of the Payne et al. (this issue) model for how people reason about implicit, relative to explicit, bias. On one hand, conceptualizing implicit bias as a product of mental associations that are triggered in situations may lead people to be more willing to accept that their judgments, decisions, and behaviors could be susceptible to the influence of implicit bias (see Uhlmann & Nosek, 2012). Research has found that learning that one does indeed harbor implicit bias, especially biases that are deemed socially unacceptable in their explicit form (e.g., pro-White racial bias), is threatening to the self-concept of many Americans, at least initially (e.g., Howell & Ratliff, 2016). A model of implicit bias that situates its expression on situational factors, then, should be more acceptable to individuals. To extend the "wave"

metaphor introduced by Payne et al. (this issue), just like we think of an individual who stands at the appropriate time to participate in the wave at a sporting event differently than someone who engages in the same behavior far too soon, too late, and especially in the absence of the wave, people are likely to think of themselves (and perhaps others) differently if implicit bias is largely considered situational rather than individual.

Further, given the importance of accepting the possibility that implicit bias exists and might shape one's own decisions for interventions on such influences (Devine, Forscher, Austin, & Cox, 2012; Holroyd, 2012) the bias of crowds model of implicit bias may ultimately result in greater support for efforts designed to combat the effects of implicit bias. Alternatively, it is possible that the proposed shift in how implicit bias is conceptualized could exacerbate what largely seems to be a general trend to hold perpetrators of implicit bias *less* accountable for discriminatory actions than perpetrators of explicit bias (Daumeyer, Onyeador, & Richeson, in preparation; see also Onyeador, 2017). Indeed, a series of studies in our lab described the biased behaviors that doctors revealed toward older adult patients or patients engaging in politicized health behaviors (e.g., marijuana use, gun ownership) as being due to either implicit (unconscious) or explicit (conscious) attitudes and being either somewhat intentional or totally unintentional. Participants then reported on the doctors' level of accountability for their behavior, as well as on their level of concern about bias in medical encounters and support for educational and other reforms to combat such biases. Findings consistently suggest that participants held perpetrators of discrimination born of implicit bias *less* accountable than perpetrators of discrimination born of explicit bias.

Interestingly, participants also tended to express equivalent levels of concern about bias in medical encounters and support for education and reforms, irrespective of whether the discriminatory behavior was framed as due to implicit or explicit attitudes. Onyeador (2017) observed similar findings for racial bias described as implicit or explicit. Indeed, White participants perceived discrimination as less intentional and, as a consequence, less harmful and held perpetrators less accountable for it if the bias was framed as implicit rather than explicit. Taken together, then, this emerging work suggests that current understandings of implicit bias have created an accountability gap and even may have reduced the perceived severity of some forms of discrimination. A model that shifts the source of implicit bias from individuals to situations, then, should exacerbate this accountability gap. If implicit bias is "a social phenomenon that passes through [individuals] rather than residing in [individuals]" (Payne et al., this issue, p. 236), it seems reasonable to expect that people will hold the perpetrators of discrimination that is attributed to implicit rather than explicit bias even less accountable than they do currently. The bias of crowds model, in other words, may make it even less likely that people will hold individual actors accountable for their discriminatory behaviors that are due to implicit bias. Although this may be an unintended consequence of the model, it would certainly benefit those concerned about social justice to anticipate this likely outcome and, perhaps, develop ways to combat it so as to hold individual actors accountable while forwarding an accurate portrait of the ways in which implicit bias actually operates.

Although we are concerned about the accountability gap regarding implicit versus explicit bias, we also believe that there are likely to be many benefits of increasing the salience of the role of situations in the emergence of implicit bias. Most notably, such a shift may lead scholars and practitioners alike to refocus their reform efforts away from individual minds and toward structural forces and practices that can disrupt the expression and influence of implicit bias in decision making. More important, however, the bias of crowds model may lead people to scrutinize persistent forms of stratification and inequality that have been normalized or explained away. For instance, if residential racial discrimination and the even more pronounced patterns of school segregation are once again perceived as *causes of* racial bias (implicit bias, in this case) rather than as rooted in such biased beliefs, it is possible that parents may rethink their decisions regarding neighborhood and school characteristics. In other words, the bias of crowds model of implicit bias may inspire lay people and researchers alike to reverse the causal arrow regarding racial (and other forms of) bias and homogenous social contexts (neighborhoods, social networks, educational institutions, employment sectors, etc.)— a topic we consider more thoroughly in the next section. Of course, such efforts may be efficacious only to the extent that people are able to discern which context and situations are affecting their thoughts and behaviors—albeit largely automatically—and, ultimately, motivated and/or able to change those contexts and situations. To the extent that such motivations and opportunities are shaped by individual differences, is there not a case to be made for an individual difference approach to implicit bias after all?

In sum, in shifting the conceptualization of implicit bias from something that individuals have to something that they experience by virtue of the situations and contexts they inhabit, the bias of crowds model is likely to affect how people reason about implicit bias, be it their own or that attributed to others. Whereas this shift may reduce the extent to which people are reluctant to acknowledge their own implicit biases, it may increase their reluctance to hold others accountable for actions that are due to implicit bias. But if individuals are not responsible for discrimination based on implicit bias, who or what is? Further, whereas renewed attention to the social contexts that promote biased mental associations is clearly paramount to gaining clarity on what implicit bias is and how it actually operates, there is also reason to suspect that shifting the focus and metrics from individual to aggregate levels of implicit bias will be summarily dismissed as unfairly condemning whole regions of the country as racist or sexist. Navigating these and other potential pitfalls of the implications of this model is essential.

(Re)Embedding Implicit Bias in Systems, Structure, and Society

As alluded to in the last section, a primary benefit of the bias of crowds model is its potential to increase attention to the ways in which our environments are currently and constantly shaping our mental associations, including our implicit biases. Indeed, some of the most promising "interventions" into

implicit bias are basically reconstitutions of stereotype-relevant contexts—or proxies of such reconstitutions. Consider, for example, the compelling work by Dasgupta and Asgari (2004) measuring the implicit gender– science, technology, engineering, and math (STEM) bias held by high-achieving female students at the beginning of their 1st year attending either a single-sex or coeducational liberal arts college and then again after one year in these two quite different contexts. At the initial testing period—just at the start of college—the level of implicit gender bias observed was actually quite minimal and did not differ between the two cohorts. After just one year in college, however, whereas the students who attended the coed college began to reveal implicit bias against women (vs. men) in STEM, their counterparts who attended the all-female college did not. Further, exposure to female leaders and professors, especially in STEM fields, predicted the extent to which participants failed to demonstrate the implicit bias. In other words, the microenvironment of the all-female, elite university—and its correlates, such as the presence of women leaders and professors—seemed to protect against the emergence of self-defeating implicit beliefs about gender and ability.

Similarly, exposure to counterstereotypical racial exemplars has been found to reduce implicit racial bias in the lab (Dasgupta & Greenwald, 2001). Although this laboratory demonstration may be fairly unstable and short-lived as an intervention given the associations with which people are typically bombarded regarding different racial groups (Lai et al., 2016), it does suggest the possibility that different contexts yield different patterns of mental associations. Indeed, more than anything else, this is likely the reason why Black Americans, on average, reveal little to no pro-White or pro-Black implicit racial bias; namely, societal forces and exposure increase pro-White associations, whereas racial socialization and the presence of positive Black exemplars in people's actual environments (and families) nudges them toward pro-Black associations (Axt, Ebersole, & Nosek, 2014; Hughes et al., 2006; Livingston, 2002). The bias of crowds model, in other words, serves to remind us that our immediate social contexts are continuously shaping the mental associations that inform our decisions and behavior and, yes, that we express on measures of implicit bias. Consequently, similar to recent theoretical work focused more directly on racial bias (Richeson & Sommers, 2016), this model calls for renewed attention to the relevance of sociocontextual and, perhaps, sociohistorical factors that continue to shape and reshape our implicit beliefs and biases.

As empirical research begins to test this model, however, it will be crucial to consider the nature of different relationships between implicit bias measurement (expression) and markers of societal or even local inequality. Consider, for instance, the compelling evidence presented in Payne et al. (this issue) that aggregate levels of implicit racial bias in a given city predict racial disparities in police shootings. Given that the majority of people measured on the implicit bias tests were not themselves the police officers committing the brutality and violence against Blacks, how should we understand this correlation and other aggregate-level correlations between implicit bias and inequality markers? Is there a third variable that is predictive of both implicit bias and discrete markers of inequality, such as broader, long-standing cues in the environment regarding contingencies about group membership, status, and value? Or is there a direct causal relationship between use of violence against Black citizens in a community by the police and the level of implicit bias held by members of the community? Or, rather, does the implicit bias reflect a tolerance of—that is, apathy regarding—police shootings? Gaining clarity regarding the ways in which different measures of inequality and implicit bias are related, especially the potential causal relations, is of paramount importance.

Similarly, although we concur with Payne et al. (this issue) that implicit biases reflect societal inequality, theoretical and empirical work is needed to interrogate how this information is transmitted across individuals in a given context to influence their implicit attitudes. It may be particularly important to uncover how people come to understand inequality within their broader social context (e.g., their city or state), especially in light of considerable evidence that our perceptions of societal inequality can deviate quite far from reality. For instance, in the case of economic inequality both in general (Norton & Ariely, 2011) and as a function of race (Kraus, Rucker, & Richeson, 2017), Americans perceive far less inequality than actually exists. Indeed, individuals are largely motivated *not* to perceive the actual degree of societal inequality in many domains. So, then, is it the actual level of inequality that shapes implicit bias or people's imagined level?

Further, although we fully agree with Payne and colleagues (this issue) that work on implicit bias (at least toward societal groups) is often a reflection of societal inequality, this should not distract from the many other more dynamic societal forces that inform implicit (and explicit) bias and intergroup relations more broadly. Consider, for instance, classic research noting the role of threats to a group's societal standing on the expression of racial bias (Blalock, 1967; Blumer, 1958; Bobo & Hutchings, 1996; see also Stephan & Stephan, 2000). These dynamics have recently been unearthed in the ways in which Whites are responding to information about the shifting racial demographics of the nation (e.g., Craig & Richeson, 2014a, 2014b; Craig, Rucker, & Richeson, in press; Danbold & Huo, 2015; Outten, Schmitt, Miller, & Garcia, 2012; Skinner & Cheadle, 2016). Similarly, relatively stable individual differences like one's degree of preference of social hierarchy (e.g., social; dominance orientation; Sidanius & Pratto, 1999) has been linked to both differences in implicit bias (Pratto & Shih, 2000) and perceptions of inequality (see Pratto, Sidanius, & Levin, 2006). In other words, future research on the bias of crowds model would benefit from thoughtful consideration of how this new conceptualization of implicit bias fits into broader sociological and psychological frameworks of intergroup hierarchy and inequality.

Conclusion

The bias of crowds model is poised to a launch a critical interrogation into the ways in which implicit biases are influenced by, and implicated in, continuing societal intergroup inequality. This is an important direction for implicit bias research and the social psychology of intergroup relations in general. Although this new conceptualization may exacerbate the tendency to discount discrimination that is attributed to implicit bias, its

promise lies in (a) offering a more accurate account of implicit bias and (b) reconnecting research on implicit bias more tightly to research on intergroup relations and inequality. Indeed, rather than looking toward implicit bias as a cause of societal inequality, perhaps this model will encourage us to think of implicit bias more like a barometer of inequality—a canary in the coal mine, so to speak, alerting us to the toxic, suffocating levels of inequality in our environments.

References

Axt, J. R., Ebersole, C. R., & Nosek, B. A. (2014). The rules of implicit evaluation by race, religion, and age. *Psychological Science*, *25*(9), 1804–1815.

Blalock, H. M., Jr. (1967). *Toward a theory of minority-group relations*. New York, NY: Capricorn Books.

Blumer, H. (1958). Race prejudice as a sense of group position. *Pacific Sociological Review*, *1*, 3–7.

Bobo, L., & Hutchings, V. L. (1996). Perceptions of racial group competition: Extending Blumer's theory of group position to a multiracial social context. *American Sociological Review*, *61*(6), 951–972.

Craig, M. A., Rucker, J. M., & Richeson, J. A. (in press). The pitfalls and promise of increasing racial diversity: Threat, contact, and race relations in the 21st century. *Current Directions in Psychological Science*.

Craig, Maureen A., & Richeson, J. A. (2014a). More diverse yet less tolerant? How the increasingly-diverse racial landscape affects White Americans' racial attitudes. *Personality and Social Psychology Bulletin*, *40*, 750–761.

Craig, M. A., & Richeson, J. A. (2014b). On the precipice of a "majority-minority" America: Perceived status threat from the racial demographic shift affects White Americans' political ideology. *Psychological Science*, *25*, 1189–1197.

Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, *82*, 359–378.

Danbold, F., & Huo, Y. J. (2015). No longer "all-American"? Whites' defensive reactions to their numerical decline. *Social Psychological and Personality Science*, *6*, 210–218.

Dasgupta, N., & Asgari, S. (2004). Seeing in believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping, *Journal of Experimental Social Psychology*, *40*, 642–658.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*, 800–814.

Daumeyer, N. M., Onyeador, I. N., & Richeson, J. A. (in preparation). Consequences of framing bias as implicit versus explicit. Manuscript in preparation.

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. L. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking interventions. *Journal of Experimental Social Psychology*, *48*(6), 1267–1278.

Holroyd, J. (2012). Responsibility for implicit bias. *Journal of Social Philosophy*, *43*, 274–306.

Howell, J. L., & Ratliff, K. A. (2016). Not your average bigot: The better-than-average effect and defensive responding to Implicit Association Test feedback. *British Journal of Social Psychology*, *56*(1), 125–145.

Hughes, D., Rodriquez, J., Smith, E. P., Johnson, D. J., Stevenson, H. C., & Spicer, P. (2006). Parents' ethnic–racial socialization practices: A review of research and directions for future study. *Developmental Psychology*, *42*(5), 747–770.

Kraus, M. W., Rucker, J. M., & Richeson, J. A. (2017). The misperception of racial economic equality. *Proceedings of the National Academy of Sciences*.

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., … Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*, *145*, 1001–1016.

Livingston R. W. (2002). The role of perceived negativity in the moderation of African Americans' implicit and explicit racial attitudes. *Journal of Experimental Social Psychology*, *38*, 405–413.

Norton, M. I., & Ariely, D. (2011). Building a better America—One wealth quintile at a time. *Perspectives on Psychological Science*, *6*, 9–12.

Onyeador, I. N. (2017). *Presumed unintentional: The ironic effects of implicit bias framing on Whites' perceptions of discrimination* (Unpublished doctoral dissertation). Los Angeles, CA: University of California.

Outten, H. R., Schmitt, M. T., Miller, D. A., & Garcia, A. L. (2012). Feeling threatened about the future: Whites' emotional reactions to anticipated ethnic demographic changes. *Personality and Social Psychology Bulletin*, *38*, 14–25.

Pratto, F., & Shih, M. (2000). Social dominance orientation and group context in implicit group prejudice. *Psychological Science*, *11*, 515–518.

Pratto, F., Sidanius, J., & Levin, S. (2006). Social dominance theory and the dynamics of intergroup relations: Taking stock and looking forward. *European Review of Social Psychology*, *17*, 271–320.

Richeson, J. A., & Sommers, S. R. (2016). Toward a social psychology of race and race relations for the twenty-first century. *Annual Review of Psychology*, *67*, 439–463.

Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. New York, NY: Cambridge University Press

Skinner, A. L., & Cheadle, J. E. (2016). The "Obama effect"? Priming contemporary racial milestones increases implicit racial bias among Whites. *Social Cognition*, *34*, 544–558.

Stephan, W. S., & Stephan, C. W. (2000). An integrated threat theory of prejudice. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 23–46). Mahwah, NJ: Erlbaum.

Uhlmann, E. L., & Nosek, B. A. (2012). My culture made me do it: Lay theories of responsibility for automatic prejudice. *Social Psychology*, *43*, 108–113.