Controlling the Influence of Stereotypes on One's Thoughts

PATRICK S. FORSCHER and PATRICIA G. DEVINE

Abstract

Research on reducing or controlling implicit bias has been characterized by a tension between the two goals of reducing lingering intergroup disparities and gaining insight into human cognition. The tension between these two goals has created two distinct research traditions, each of which is characterized by different research questions, methods, and ultimate goals. We argue that the divisions between these research traditions are more apparent than real and that the two research traditions could be synergistic. We attempt to integrate the two traditions by arguing that implicit bias, and the disparities it is presumed to cause, is a public health problem. On the basis of this perspective, we identify shortcomings in our current knowledge of controlling implicit bias and provide a set of recommendations for future research.

CONTROLLING IMPLICIT BIAS: INSIGHTS FROM A PUBLIC HEALTH PERSPECTIVE

Within the past 15 years, there has been an explosion of research on controlling automatic stereotypes, or more generally on controlling so-called implicit biases. To people interested in improving the lives of minorities, the source of this interest is obvious—implicit biases are presumed to lead to subtle forms of discrimination, which, in turn, are assumed to lead to poor outcomes for minority groups. However, to people interested in the inner workings of the human mind, the source of this interest, while different from the source identified earlier, is equally obvious—implicit biases provide a convenient arena to glean knowledge about how people regulate their thoughts and behavior.

The tension between interest in solving a broad societal problem and interest in gaining insight into the human mind has led to two distinct research traditions on controlling implicit bias. On the one hand, a group of people from a broad array of fields, from sociology, to political science, to psychology, and with a broad range of formal research training, have

focused on controlling subtle biases primarily as a means to solve large-scale social disparities. People following this research tradition are fundamentally interested in intervention; that is, uncovering methods to reduce implicit biases. Moreover, although the disparities-focused researchers are interested in reducing implicit biases, these researchers see the reduction of implicit bias as not an end unto itself, but as merely a means to the ultimate end of making society a fairer place. Thus, the focus for these researchers is not so much on the mind itself, but rather on how knowledge of the mind might provide an anchor for understanding, and, eventually, alleviating lingering social disparities.

On the other hand, a separate group of people, most of whom are academic research psychologists, have been drawn to the field of implicit bias because it provides an interesting and important theoretical context that can be harnessed to gain insights into human cognition. A person attempting to control the influence of implicit biases must deploy the various tools of cognitive control to inhibit the influence of a set of fast, efficient cognitive processes on ongoing behavior. Thus, studying the control of implicit biases can give researchers theoretical leverage to better understand the development of automatic processes, the activation of these processes, the effects of these processes on behavior, and the ways in which cognitive control can be strategically deployed to counteract these processes. In contrast to disparities-focused researchers, researchers following the cognitive tradition take a decidedly internal, mechanistic focus; instead of orienting themselves toward a particular social problem, these researchers orient themselves inward, toward the underlying cognitive and neural mechanisms presumed to underlie the activation and subsequent control of automatic biases.

Although some research draws inspiration from both research traditions, the distinct goals of resolving societal problems and gaining insight into the human mind have created a gulf that divides the established research along both theoretical and methodological lines. Indeed, the tension created by these separate research traditions is evident in a qualitative review of 985 research reports on reducing intergroup bias conducted in 2009 by Paluck and Green. Paluck and Green reviewed a broad swath of reports, spanning both published and unpublished research studying the reduction of both implicit and explicit outcomes. They found that research on intergroup bias divided sharply along theoretical and applied lines, and this division was accompanied by differences in research questions, research quality, and research method. The applied research tended to focus on resolving real-world disparities. This research focused on mostly explicit and behavioral outcomes and tended to be conducted in real-world settings. Unfortunately, the applied research also tended to use designs that did not permit sound causal inference; fully 60% (581) of the total reviewed studies

were nonexperimental, of which only 38% used a control group. In contrast, theoretically oriented research tended to focus on implicit outcomes. While this research often used randomized controlled designs, it was also most often conducted in artificial laboratory settings; of the total 391 experimental studies reviewed, only 107 were conducted in the field.

Paluck and Green's review speaks to the deep divisions between disparities-focused research and cognition-focused research. Although these two research traditions do indeed seem to be divided by differences in questions, methods, and goals, we believe that these differences are more apparent than real. Indeed, by focusing on common theoretical questions and by harnessing their respective research strengths, we believe that these two approaches can be synergistic. By working together, researchers following the cognitive and disparities traditions can bring their respective strengths to bear on real-world problems in ways that substantively advance our knowledge of how we can change individual minds to resolve societal-level problems.

In what follows, we review the past and present research on controlling automatic bias with a view toward uniting the cognitive-focused and disparities-focused research traditions. We attempt to bring these two research traditions together by borrowing a perspective that has been successful in integrating theory and practice in other fields—a public health perspective. As we describe later, the public health perspective provides incentives for the accumulation of knowledge at multiple levels of analysis by focusing on the ultimate goal of improving the health of larger populations. Focusing on the goal of improving public health also highlights gaps in our knowledge by orienting us toward the specific steps needed to make substantive improvement in public health. After reviewing past and current research, we use the public health perspective to provide a set of recommendations for future research.

CLASSIC RESEARCH: A FOCUS ON A SPECIFIC SOCIAL PROBLEM

In order to understand the tensions between research traditions in implicit bias research, it is helpful to consider the historical context from which the field arose. At the close of the Civil Rights Movement, civil rights activists had achieved a fundamental change in the formal and informal norms governing intergroup relations. Formally, overt discrimination had been outlawed. Informally, overt discrimination had become socially taboo. Violations of these formal and informal norms were subject to severe economic, legal, and social sanctions.

Overt discrimination had thus been made extremely difficult to perpetrate, at least in the presence of a disapproving audience. However, the hope

among activists was that the changes in overt discrimination would generalize to changes in covert discrimination. The logic behind this hope was that external pressures would instigate changes within individual people by encouraging people to monitor their own behavior, regardless of whether external audiences were also monitoring their behavior. These internalized monitoring processes would prevent people from discriminating against outgroups even when the discrimination could not be punished by others. In essence, activists hoped that external pressure could create internal change in people's underlying psychology.

Despite these hopes, disparities between Black people and White people[1] linger across a wide variety of domains, from educational attainment, to economic success, to overall health, to psychological well-being (e.g., Bertrand & Mullainathan, 2004; Steele, 1997). Moreover, the disparities do not seem to be perpetuated by a few ill-intentioned people; even people who report that they believe prejudice is wrong seem, paradoxically, to discriminate against Black people in subtle ways (Crosby, Bromley, & Saxe, 1980; Devine, 1989). Because the source of the continuing disparities endangers the health and well-being of an entire population, the cause or causes of these disparities constitute a broad-scale public health crisis. The task for someone concerned about eliminating the disparities is to identify these causes to discover whether they are amenable to change, and if so, to pursue strategies to reduce them.

Unfortunately, two problems stand in the way of identifying and acting upon the causes of racial disparities. The first of these is the *targeting problem*, which refers to the issue of deciding the points at which intervention can successfully alleviate lingering disparities. The targeting problem exists because of an enduring ambiguity about the precise mechanisms through which racial disparities perpetuate themselves. To the extent that we know the causes of racial disparities, we should be better able to devise ways to act upon those causes to alleviate the disparities.

The second problem is the *measurement problem*, which refers to the difficulty of defining and quantifying progress toward the goal of reducing racial disparities. Although the solution to this problem might seem obvious—why not simply measure the disparities themselves?—at least two issues create obstacles for this solution. First, disparities between majority and minority group members exist across a broad swath of domains, and without a clear measurement benchmark that is applicable across these domains, it is unclear how to integrate knowledge acquired about interventions in one domain

---

1. Although disparities exist between many social groups, and although implicit bias may play a role in many of these disparities, in this essay we focus primarily on disparities between Black people and White people because of their historical importance to the social and political movements that influenced implicit bias research in the United States.

with knowledge acquired about interventions in another domain. Second, and perhaps even more importantly, disparities exist at a structural level and are likely sustained and perpetuated by a large number of causal factors. Therefore, it is possible for a small-scale intervention to advance the overall goal of eliminating disparities by eliminating one of the causal factors, but not result in change in the overall disparities because of the existence of the other causal factors sustaining the disparities.

One response to the targeting and measurement problems, and a response favored by psychologists, is to simplify the problem by focusing on an individual level of analysis. The assumption behind this approach is that if we are able to develop effective interventions that create change within individuals, we can then deploy these interventions on a large enough scale to create change on a societal level. The task of eliminating disparities thus can be simplified into the task of identifying the source of the individual-level paradox of why well-intentioned people nonetheless continue to discriminate against outgroups.

Thus far, our review has mainly followed researchers inspired by the more disparities-focused tradition of studying implicit bias. Here, however, the paths of the disparities-focused researchers converge with those of the cognitive researchers. In the early 1980s and 1990s, the cognitive revolution had fully infused social psychology, and social psychologists had begun to use the tools of the cognitive revolution to ask new questions about the processes underlying social phenomena and to provide measures of those processes. The cognitive revolution brought with it a unique analysis of social behavior—the idea that processes that lead to behavior can become automatized to the point where they no longer require conscious activation, and that these processes can lead to behaviors that are neither intended nor desired (Devine, 1989). Alongside this analysis came measures that relied on priming and the measurement of reaction times that could be used to probe and investigate these automatic processes.

Researchers following the social cognition tradition appeared to provide the theoretical and methodological tools necessary to solve conundrums posed by the targeting and measurement problems. Their theoretical analysis provided an understanding of how individual people could be unwittingly complicit in the perpetuation of racial disparities despite intentions to act fairly, while their methodological tools provided a means to measure the processes implicated in unintentional discrimination (e.g., Greenwald, McGhee, & Schwartz, 1998). By providing the tools to measure a cognitive process assumed to play a causal role in the perpetuation of disparities, social cognitive researchers provided a benchmark against which progress toward the goal of reducing racial disparities could be assessed. This common benchmark could then be used to integrate knowledge about the

relative effectiveness of a broad range of interventions, thereby facilitating the advance of knowledge across a broad range of fields toward solving a large-scale public health issue.

CURRENT RESEARCH: A FRACTURED FIELD

As already described above, the challenges involved in identifying the causes of racial disparities and identifying ways to act upon those causes is extremely complex. It has taken a considerable amount of time for scholars to reach the tentative conclusion that implicit bias might be one of these causes, with the result that the study of interventions to control implicit bias is still in its infancy. From this perspective, it is therefore not surprising that we currently have only limited knowledge of effective interventions to change both implicit bias and the disparities presumed to be caused by implicit bias.

However, even accounting for the youthful state of the field, current research on interventions to change implicit bias is fractured. Although focusing on implicit bias and measures of implicit bias has given researchers a common benchmark with which to judge the effectiveness of various interventions, researchers following the disparities-focused and cognitive research traditions have continued to pursue separate research goals, resulting in a literature that is rather scattered and difficult to interpret.

Social cognitive researchers have focused primarily on either advancing or challenging theories of human cognition, often by implicating specific cognitive mechanisms behind an experimental effect. This research, while useful for shedding light on cognitive mechanisms, has not always advanced our understanding of how to reduce lingering racial disparities. For example, a voluminous literature has developed regarding the malleability of implicit bias. This literature arose primarily as a reaction to theoretical portrayals of implicit bias as inevitable and immutable (e.g., Bargh 1999), and thus, the focus of this literature is in providing demonstrations that bias on implicit measures was amenable to change. Although the malleability literature has provided convincing demonstrations that responses on implicit measures can be changed and has even uncovered some of the mechanisms behind these changes (e.g., Payne, 2001), little of this research has gone on to show that these changes on implicit measures are consequential in that they last over time and generalize to consequential behaviors.

In another example, Phills, Santelli, Kawakami, Struthers, and Higgins (2011) argued that messages presented against background stimuli that are concordant with those messages are more effective in reducing implicit bias than messages presented against discordant backgrounds. Such a result would be interpreted as evidence supporting regulatory focus theory

(Higgins, 1997), a theory of self-regulation. Accordingly, Phills *et al*. found that presenting the message, "Say yes to equality," which has an approach orientation, is more effective when presented with background pictures of positive interracial interactions than when presented against background pictures of the KKK. Although these findings are interesting, they advance our understanding of regulatory focus theory more than they advance our understanding of how to reduce lingering racial disparities. Overall, researchers following the social cognitive tradition have focused more on advancing our understanding of human cognition rather than on advancing our understanding of how to resolve the social problems presumed to be caused by implicit bias.

Researchers focusing more on social disparities have, for their part, either sought to demonstrate that implicit bias is related to negative intergroup outcomes at an interpersonal level or sought "natural experiments" that suggest routes through which implicit bias might be reduced. As an example of the first category of research, Richeson and Shelton have conducted a series of studies showing how the implicit bias of Whites is related to interaction quality for both White and Black participants (e.g., Richeson & Shelton, 2003). Although this research is important for establishing the validity of measures of implicit bias, it provides a mere snapshot of one outcome and does not situate that outcome in a broader social context.

As an example of the second type of research, Rudman, Ashmore, and Gary (2001) investigated whether people who enrolled in a class on modern racism had lower levels of implicit bias at the end of the academic semester than people who enrolled in a research methods class. By examining both implicit and explicit outcomes, the researchers hoped to find evidence supporting the argument that implicit biases can be overcome through intensive, long-term experience. Although this research has the considerable advantage of examining implicit outcomes over time, it also does not address whether the changes observed are related to changes in behaviors that contribute to racial disparities, and because the design is correlational, the precise interpretation of why the changes occurred is ambiguous.

Overall, the fractured status of the field has had the unfortunate result that although a number of promising interventions have been identified by social cognitive researchers, we have little knowledge of whether the effects of those interventions are consequential. Moreover, we have little theoretical understanding of how the various interventions identified by these researchers relate to each other. On the other hand, from the disparities-focused researchers, we have little knowledge of whether the factors that are related to decreased implicit bias are causally related to reduced disparities.

Despite these limitations, there are a few promising trends that suggest that more researchers are starting to attend to ensuring that their interventions are related to meaningful change in implicit bias. For example, Devine, Forscher, Austin, and Cox (2012) examined the effects of a randomized, multifaceted training intervention on implicit and explicit outcomes over the course of 2 months. They found that their intervention was related to reduced implicit bias up to 2 months after the administration of the intervention. Although this study did not examine whether the reductions in implicit bias were accompanied by reductions in the behaviors that contribute to lingering disparities, it does attempt to integrate the controlled, mechanistic focus of social cognition researchers with the goal of reducing disparities by couching the reduction of long-term personal bias in terms of societal-level problems. Our hope is that future research can follow the example set by Devine and colleagues by showing how the effects of their interventions create meaningful change.

FUTURE RESEARCH: RECOMMENDATIONS FROM A PUBLIC HEALTH PERSPECTIVE

From a public health perspective, the major limitations in our current understanding of reducing implicit bias relate to the fractured state of the field and the fact that we have not effectively connected implicit outcomes to behaviors contributing to disparities. Integrating the field through collaboration between researchers from the social cognition and disparities traditions would go far in creating a more integrated field. These collaborative efforts would also help stimulate creative research that permits sound causal inference while utilizing the realistic, real-world settings where disparities occur. Thus, one of our primary recommendations is that researchers from the social cognition and disparities traditions reach out to each other to produce synergistic research that can more completely address the challenges in both research traditions.

In addition, to the extent that implicit bias does constitute a substantive public health problem, we believe that implicit bias researchers can benefit by borrowing strategies adopted by other fields concerned with public health. To address the problem of tying implicit bias to the disparities presumed to be caused by implicit bias, we believe we can borrow strategies adopted by epidemiologists. Epidemiologists specialize in untangling the patterns, causes, and effects of disease and other health conditions within populations, and to this end, employ a variety of descriptive, observational, and experimental methods to aid in their understanding. In a similar way, scholars who view lingering disparities as a public health problem and who suspect that implicit bias is one of its causes should attend to the patterns, causes, and

effects of implicit bias within populations, with an eye toward understanding the factors that strengthen and weaken the relationship between implicit bias and disparities at a population level. This epidemiological work will go far toward advancing the goal of understanding the extent to which implicit bias plays a causal role in perpetuating lingering disparities.

To the extent that implicit bias is causally related to lingering disparities, researchers wishing to reduce these disparities could borrow from clinical trials research. Because the intent of clinical trials is to discover interventions that improve overall public health, researchers who implement clinical trials typically emphasize open access, the ability to make sound causal inference, the use of a broad range of participant populations, replicability, and the discovery of interventions that produce lasting, meaningful outcomes. Thus, clinical trials are registered on a public website (clinicaltrials.gov), utilize double-blind designs that often compare multiple interventions within the same study, measure the effects of their interventions over time, and, when particularly promising interventions are tested, use large sample sizes. Although many implicit bias researchers are already cognizant of the importance of some of these tenets, we do not always implement these tenets in our research methods.

Finally, to integrate the findings from clinical trials into a coherent body of evidence, public health researchers emphasize the importance of frequent, high-quality meta-analysis. Meta-analysis is widely acknowledged to be a crucial contributor to the accumulation of scientific knowledge, and yet, to date, no one has meta-analytically evaluated the effectiveness of implicit bias interventions. Although forthcoming work may address this shortcoming (Lai *et al.*, in preparation), regular meta-analysis would better inform researchers of the current state of the art and would enable implicit bias research to become a more cumulative field.

Overall, the goal of this essay was to highlight the differences between the cognitive and disparities-focused research traditions and how the separation between these two traditions has contributed to a fractured research literature. We suggest that these two traditions can be united by taking a public health perspective on social disparities. A public health perspective focuses researchers on the underlying societal problem that originally stimulated interest in implicit bias and emphasizes careful description of that problem, the identification of the causes of the problem, and the accumulation and integration of knowledge about that problem. Although the public health perspective departs from the customary perspectives of psychologists focused on individual-level analyses, the perspective highlights assumptions inherent in more customary perspectives and forces us to delineate the ways in which our research matters.

REFERENCES

Bargh, J. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 361–382). New York, NY: Guilford Press.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, *94*, 991–1013. doi:10.1257/0002828042002561

Crosby, F., Bromley, S., & Saxe, L. (1980). Recent unobtrusive studies of Black and White discrimination and prejudice: A literature review. *Psychological Bulletin*, *87*(3), 546–563. doi:10.1037/0033-2909.87.3.546

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. doi:10.1037/0022-3514.56.1.5

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 intervention. *Journal of Experimental Social Psychology*. doi:10.1016/j.jesp.2012.06.003

Lai, C. K., Forscher, P. S., Ebersole, C., Axt, J., Herman, M., Devine, P. G., & Nosek, B. A. (in preparation). A meta-analysis of interventions to change implicit bias.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480.

Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist*, *52*(12), 1280–1300. doi:10.1037/0003-066X.52.12.1280

Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? A review and assessment of research and practice. *Annual Review of Psychology*, *60*, 339–367. doi:10.1146/annurev.psych.60.110707.163607

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*(2), 181–192. doi:10.1037/0022-3514.81.2.181

Phills, C. E., Santelli, A. G., Kawakami, K., Struthers, C. W., & Higgins, E. T. (2011). Reducing implicit prejudice: Matching approach/avoidance strategies to contextual valence and regulatory focus. *Journal of Experimental Social Psychology*, *47*(5), 968–973. doi:10.1016/j.jesp.2011.03.013

Richeson, J. A., & Shelton, J. N. (2003). When prejudice does not pay: Effects of interracial contact on executive function. *Psychological Science*, *14*(3), 287–290. doi:10.1111/1467-9280.03437

Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). "Unlearning" automatic biases: the malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology*, *81*(5), 856–868. doi:10.1037/0022-3514.81.5.856

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613–629. doi:10.1037/0003-066X.52.6.613

FURTHER READING

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. doi:10.1037/0022-3514.56.1.5

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 intervention. *Journal of Experimental Social Psychology*. doi:10.1016/j.jesp.2012.06.003

Forscher, P. S., & Devine, P. G. (2014). Breaking the prejudice habit: Automaticity and control in the context of a long-term goal. In J. Sherman, B. Gawrsonski & Y. Trope (Eds.), *Dual process theories of the social mind*. New York, NY: Guilford Press.

Lai, C. K., Forscher, P. S., Ebersole, C., Axt, J., Herman, M., Devine, P. G., & Nosek, B. A. (in preparation). A meta-analysis of interventions to change implicit bias.

Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? A review and assessment of research and practice. *Annual Review of Psychology*, *60*, 339–367. doi:10.1146/annurev.psych.60.110707.163607

PATRICK FORSCHER SHORT BIOGRAPHY

Patrick Forscher is a Psychology graduate student at the University of Wisconsin-Madison. He is deeply interested in intervention and convinced of the necessity of putting our research to work in society. In his spare time, Patrick enjoys that quintessentially Wisconsin hobby, homebrewing.

PATRICIA G. DEVINE SHORT BIOGRAPHY

Patricia G. Devine is Professor of Psychology at the University of Wisconsin-Madison. An experimental social psychologist who specializes in prejudice, stereotypes, and intergroup relations, she is interested in how people manage the intrapersonal and interpersonal challenges associated with prejudice in our contemporary society. One main focus for recent work has been the sources of motivation, internal and external, for responding without prejudice, and the unique challenges these alternate sources of motivation create for managing the interpersonal aspects of intergroup relations.

RELATED ESSAYS

Empathy Gaps between Helpers and Help-Seekers: Implications for Cooperation *(Psychology)*, Vanessa K. Bohns and Francis J. Flynn

Identity Fusion *(Psychology)*, Michael D. Burhmester and William B. Swann Jr.

Mental Models *(Psychology)*, Ruth M.J. Byrne