Psychological Science Faculty Publications and
Presentations

Psychological Science

5-17-2017

Knowledge-based interventions are more likely to reduce legal disparities than are implicit bias interventions

Patrick S. Forscher
University of Arkansas, Fayetteville, forscher@uark.edu

Patricia G. Devine
University of Wisconsin-Madison

Knowledge-based interventions are more likely to reduce legal disparities than are implicit bias interventions

Patrick S. Forscher and Patricia G. Devine

A preprint version of a chapter written for the book

Enhancing Justice: Reducing Bias
Redfield, S. E., & American Bar Association (Editors), 2017

Scholars have long noted lingering disparities in United States legal outcomes between social groups (Mustard, 2001; Gross & Mauro, 1989). These disparities are troubling by themselves because they suggest that some groups are treated unfairly before the law. However, they are especially concerning given that judges and other legal decision-makers have a duty to uphold and interpret the law in a way that is neutral with respect to age, race, gender, and other protected categories. If legal actors sincerely intend to uphold their legal duties, the existence of ongoing disparities implies that forces other than intentions contribute to ongoing disparities.

One factor in particular, implicit bias, has received increasing levels of attention. The existence of this volume attests to the fact that legal scholars are both interested in and concerned about how implicit bias may undermine fairness in the judgments of those charged with adjudicating evidence in the legal setting. Defined as automatically activated associations about social groups (Forscher, Lai, et al., under review), implicit biases are theorized to be acquired early in life through repeated exposure to stereotypic information (Devine, 1989). Over time, implicit biases become so well-rehearsed that, upon exposure to information related to a stereotyped group, they activate automatically and without intention. Once activated, these biases are theorized to affect behavior in situations where people lack motivation, awareness, or the ability to think about their responses (Devine, 1989; Fazio & Olson, 2014).

Across a large number of studies, scholars have observed that scores on measures of implicit bias are associated with behavior across a large number of domains (Greenwald et al., 2009; Cameron et al., 2012; Oswald et al., 2013). Based on this evidence, many have argued that implicit bias causes biased behavior, and that the accumulation of these biased behaviors causes disparities (Greenwald, Banaji, & Nosek, 2014; Jost et al., 2009). The mere existence of disparities in legal outcomes is troubling. However, if it is true that implicit bias causes disparities, the existence of implicit bias raises a possibility that is even more disquieting: well-meaning legal actors who sincerely desire to uphold their duties to the law may nonetheless be complicit in perpetuating these disparities. It is therefore important to evaluate the evidence as to the likely cause of legal disparities, both to determine whether this disquieting possibility has been realized and to evaluate what we can do to alleviate disparities in legal outcomes. Evaluating the cause of outcomes is challenging, but it is especially challenging when the cause is theorized operate inside people's heads without their conscious awareness. This task is the province of many social scientists, but especially psychological scientists.

We should note at the outset that this chapter is different from most others in this volume. Neither author is an expert of the law, legal proceedings, or the criminal justice system more generally. Instead, we are both psychological scientists who specialize in race and unintentional forms of bias. Our goal in this chapter is to review the extant work on implicit bias and interventions to change implicit bias. Though the work in this area is ever burgeoning, the evidence regarding the effectiveness of implicit bias interventions is rather mixed and the goals for the specific research efforts are quite varied. Our hope is to lay out the issues in ways that elucidate what this work can and cannot say about the causes of and remedies to disparities in legal outcomes. We hope that this review will prove helpful to those who choose to investigate the impact of biases, both unintentional and intentional, throughout the legal system.

To wit, in this chapter, we will review the evidence that, in general, is required to establish that implicit bias causes legal disparities. A great deal of the evidence regarding implicit bias and legal disparities is correlational and, of course, evidence of correlations between measured implicit bias and biased behavior is insufficient to establish cause. We will

then conduct a selective review of implicit bias intervention research and argue that this research also does not speak directly to whether interventions to change implicit bias are effective ways to resolve social disparities. Indeed, we will argue that, based on the extant evidence, any intervention that is focused on directly changing scores on measures of implicit bias is unlikely to be effective at changing social disparities. Finally, we will review an approach that we believe is more effective: change people's knowledge about how the structure of the social environment makes them complicit in the perpetuation of bias. We will illustrate this approach with an intervention that we have developed, the prejudice habit-breaking intervention.

Implicit bias as a cause of disparities

Developing interventions that are effective at reducing disparities requires first identifying and defining the processes that are thought to cause disparities. Any concept without a clear definition becomes hopelessly broad or ambiguous, and therefore not useful. In the context of interventions, if the definition of a process that is thought to cause disparities is unclear, it becomes difficult or impossible to craft interventions that might change that process. For these reasons, we are defining in this chapter "implicit bias" precisely as associations between concepts (for example, an association between Black people and negative concepts) that are activated automatically (i.e., without requiring awareness or intention). The measurement of implicit bias therefore requires a procedure that does not require a person to actively retrieve the target association from memory.

There are two major implications of this definition. First, implicit bias is a process that occurs within a person's head. That means that implicit bias is distinct from the disparities it is presumed to cause. Thus, the mere presence of disparities, even disparities in judgments among people, like judges, who believe in fairness, cannot by itself be taken as evidence that implicit bias causes the disparities. Implicit bias may well cause these disparities, but this must be established through an independent measurement of implicit bias. Second, implicit bias is just one type of bias that can occur unintentionally. We would like to underscore this point because it is often overlooked. Implicit bias has received perhaps more than the lion's share of focus in recent years because many prominent scholars have argued that it could play a strong role in perpetuating social disparities (Greenwald, Banaji, & Nosek, 2014; Fiske, 1998). However, many other biases can occur unintentionally without occurring because of the activation of an association between concepts. For example, legal actors can adhere to established laws and procedures, even though they themselves believe that the laws have negative consequences that disproportionately affect certain groups of people. We will revisit these points at the end of this chapter.

After a potential target of change is identified, it is essential to establish that the process actually causes disparities. Providing evidence that firmly establishes cause is quite difficult and is the subject of vigorous academic debate (Bullock, Green, & Ha, 2010; Spencer, Zanna, & Fong, 2005). However, most scholars agree that one must develop a procedure that, in randomized experiments, produces a difference in both the candidate process (i.e., implicit bias) and a behavior (i.e., discrimination). With this procedure in hand, one must show that it is the changes in the candidate process, not changes in other processes brought about by the candidate procedure, that produce the changes in the behavior. This is a challenging task.

What should be clear from this discussion is that mere associations between a process and behavior are insufficient to show that the process causes the behavior. If an intervention targets

a process that is associated with disparities but does not cause them, the result will be wasted time, effort, and, possibly, other unintended negative consequences. To illustrate why this is important, consider the following, somewhat whimsical, example. When developing policies to reduce violent crime, one might notice that the number of murders in a city is associated with the city's sales of ice cream, as both occur more frequently in larger cities. In this example, although an association exists, no one would suggest that outlawing ice cream is an effective means of reducing violent crime.

The suggestion that a correlation between violent crime and ice cream sales is clearly different from the suggestion that implicit bias causes discrimination (and therefore disparities) in that the latter suggestion is much more plausible than the former. Yet the very plausibility of this proposition makes it all the more important to evaluate it based on evidence rather than intuition. If we use intuition rather than evidence, we run the risk of committing resources to interventions that, though well-meaning, may be at best wasteful and at worst harmful. For this reason, studies that find an association between measured implicit bias and discrimination, while suggestive, should not be taken as evidence that implicit bias causes discrimination. This is true regardless of the strength of the association and regardless of the size of the sample in which the association is found (see Greenwald et al., 2009; Cameron et al., 2012; Oswald et al., 2013). Rather, to understand whether implicit bias causes disparities, we must turn to evidence based on interventions that produce changes in implicit bias and evaluate whether these changes in implicit bias bring about changes in behavior.

Implicit bias interventions

Fortunately, the extant research on interventions to change implicit bias is voluminous. We can therefore use this literature to assess whether it is possible to change implicit bias and, if so, whether these changes are likely to bring about changes in biased behavior.

Implicit bias intervention research arose out of two separate research traditions (Forscher & Devine, 2015). On the one hand are people from a broad array of backgrounds and a broad range of formal research training who are interested in implicit bias as a means of solving social disparities. On the other hand are people, mostly experimental psychologists, who are interested in implicit bias as a means of understanding the inner workings of the human mind. Some background on these research traditions is necessary to understand what they *have* and *have not* taught us about the role of implicit bias in creating social disparities.

The distinct research traditions have led to a division in both methods and research questions (Paluck & Green, 2009; Forscher & Devine, 2015). The disparities-focused researchers often conduct their research in the field, use non-experimental designs, and measure outcomes longitudinally (Paluck & Green, 2009). In addition, disparities-focused researchers do not typically measure implicit bias directly, instead focusing on self-reported or outcomes collected as part of the administrative mission of a given organization (Paluck & Green, 2009). The focus in this research is in deploying methods of solving the problems that are presumed to be caused by implicit bias while respecting the difficulties of working in non-laboratory settings.

For example, Tan, Morris, and Romero (1996) developed a training module intended to educate employees of the Federal Aviation Administration about issues related to workforce diversity, prevent workforce harassment and discrimination, and increase future FAA workforce diversity. Whether the researchers thought they were producing their effects through changes in implicit bias is unclear, though the workshop shares many features in common with other

trainings developed in the disparities-focused research tradition: it explored the primary dimensions of diversity, the values of the employees and how they might conflict with their prejudices and preconceptions, and the readiness of the employees to embrace diversity. The authors evaluated the effectiveness of the diversity training using a single-group, before-after design by asking the workshop participants to rate their knowledge of diversity-related issues. Although a single-group design respects the fact that, as a matter of organizational policy, the diversity training should not be withheld from any employees, a single-group design is not well-suited for drawing conclusions about the causes of an intervention because natural fluctuations in outcomes over time are confounded with the intervention's true effects. In addition, although self-reported knowledge is easy to measure, it is difficult for people to directly assess changes that have occurred due to a given experience (Nisbett & Wilson, 1977). Any apparent improvements in self-reported knowledge may well be due to participants wishing to present themselves as good employees who are engaged with the organizational mission of the FAA.

In contrast, cognition-focused researchers most often conduct their research in the laboratory in single-session experiments, usually using samples of college students (Forscher, Lai, et al., under review), who, though they may have psychological characteristics that make them different from the general population (Henrich, Heine, & Norenzayan, 2010), are cheap, easy, and convenient for psychological research. Cognition-focused researchers employ a large array of implicit bias measures, but measure explicit bias and behavior less often (Forscher, Lai, et al., under review). The focus of cognition-focused researchers is in drawing rigorous conclusions about mental processes.

For example, Phills and colleagues (2011) argued that messages presented on backgrounds consistent with the message's content are more effective in reducing implicit bias than messages presented on inconsistent backgrounds. Accordingly, they compared the implicit bias of people who viewed the message "Say yes to equality" presented alongside images of positive interracial interactions (a consistent pairing of message with background) to the implicit bias of people who viewed the same message presented alongside pictures of the KKK (an inconsistent pairing of message with background). They also compared the implicit bias of people who viewed the message "Say no to prejudice" presented alongside pictures of the KKK (consistent pairing) to the implicit bias of people who viewed the same message presented alongside pictures of positive interracial interactions. As predicted, Phills and colleagues found that consistent pairings produced lower scores on implicit race bias (as measured using an Implicit Association Test; Greenwald, McGhee, & Schwartz, 1998) than inconsistent pairings, and interpreted this result as supporting regulatory focus theory (Higgins, 1997), a theory of motivation. Because participants were randomly assigned to condition and because the researchers included an implicit bias measure, this study is well-positioned to speak to the causal effects of the manipulation on people's responses on that measure. However, because the researchers did not measure discriminatory behavior, the study does not speak strongly to how to reduce racial disparities. Nor was the study necessarily intended to speak to these issues—the researchers were more interested in revealing insights into the principles of motivation than they were in understanding the causes of and remedies for racial disparities.

The result of this division in method and research question is that much of the research on implicit bias interventions does not speak directly to whether these interventions are effective in resolving social disparities. Because the disparities-focused research is often not experimental, it is silent as to whether the interventions tested cause any change in social disparities. Because

implicit bias is often not measured, it is unclear whether these interventions change implicit bias at all, and, if these interventions change outcomes other than implicit bias, it is unclear whether these changes occur *because of* a change in implicit bias. The interventions developed and tested in the cognition-focused tradition usually do use randomized designs and therefore do not have the same causal inference problems as those developed and tested in the disparities-focused tradition. However, these studies have their own limitations: because they are often conducted among college students, it is unclear whether these interventions would be useful in non-college settings. Because these studies often do not include measures of behavior and often do not test their effects longitudinally, it is also unclear whether these interventions have effects on implicit bias that last and that cause changes in behavior.

Indeed, if we focus more closely on the cognition-focused studies that do measure their outcomes longitudinally, the evidence suggests that the effects of most implicit bias interventions on measures of implicit bias do not endure. For example, in a large-scale evaluation of 9 interventions to change measured implicit bias involving over 6,000 participants, all the interventions had immediate effects on implicit bias (Lai et al., 2016). However, without exception, these effects disappeared 1-2 days later. Even motivating participants to use these interventions on their own rather than at the request of the experimenter appears to be ineffective; in an evaluation of this approach with over 1,500 participants, it had no enduring effect on measured implicit bias (Forscher, Cox, et al., in preparation).

If we focus instead on the cognition-focused studies that measure both implicit and behavioral outcomes, the effects of implicit bias interventions on behavior are quite different, in general, from their effects on implicit bias. In addition, based on an analysis of more than 400 studies of methods to change implicit bias, there is no evidence that the changes in implicit bias that these interventions produce are linked to their (small) effects on behavior. That is, there is no evidence that these interventions change behavior *because of* their changes in implicit bias (Forscher, Lai, et al., under review).

Our review of research on implicit bias interventions is quite revealing. Because of issues with the design of the research (i.e., a lack of randomization, no measurement of implicit bias, no measurement of behavior), many studies that investigate implicit bias interventions do not speak strongly to whether implicit bias causes disparities. If we focus on the studies that speak more strongly to causality, the effects of current interventions to change implicit bias fade quickly, and there is no evidence that these changes in implicit bias produce changes in behavior. In sum, it is unclear whether developing interventions that change implicit bias is a reasonable approach to resolving social disparities, in law or any other domain.

A more effective approach: Knowledge-based interventions

From the perspective of classic social psychological theory, it is not surprising that interventions that directly target implicit bias appear to be ineffective. Classic social psychological theory holds that each person's conception of him- or herself strongly affects whether psychological processes are easy or difficult to change (Rokeach, 1973). According to these theories, psychological processes are organized hierarchically around a person's conceptions of themselves. Processes that are more central to the self-concept resist outside change because changing these processes requires changing all the processes that are below them in the hierarchy as well. Processes that are more peripherally located are easier to change, but these changes may not last because the processes higher in the hierarchy remain unchanged.
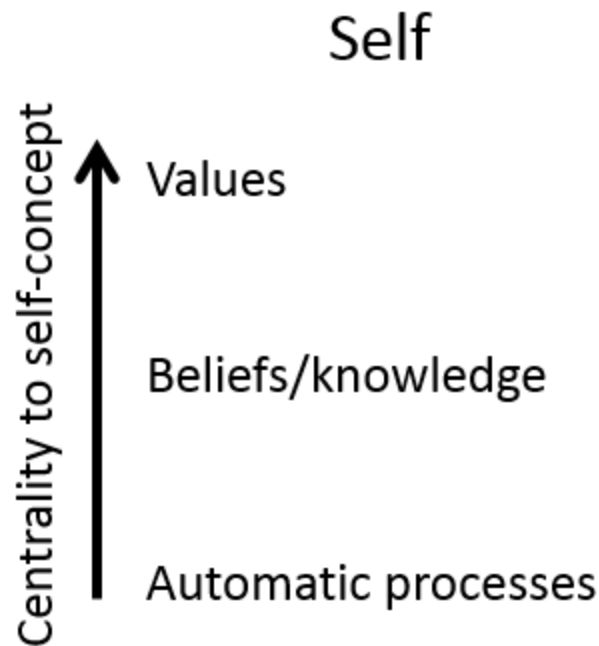
Self

Centrality to self-concept

Values

Beliefs/knowledge

Automatic processes

Figure 1. A simplified model of the self. Psychological processes that are higher in the hierarchy are more central to the self-concept and therefore more resistant to external change. Processes lower in the hierarchy are less central and therefore more susceptible to change.

Because it is a process that operates independent of awareness and intention, implicit bias a process that is by definition peripheral to the self-concept. This peripheral status implies that brief shifts in implicit bias should be easy to achieve, at least in the short term. However, implicit biases are created and sustained by repeated stereotypic pairings in the social environment. Thus, without changes in processes that are more central to the self-concept, the shifts in implicit bias are likely to be quickly erased soon after the intervention is administered (Lai et al., 2016). If the original shifts in implicit bias do not endure, they also would not be able to produce enduring changes in behavior (Forscher, Mitamura, et al., under review).

If changing implicit bias is ineffective for reducing social disparities because implicit bias is too peripheral to the self-concept, perhaps changing processes that are more central will be correspondingly more effective. However, the process that is the target of intervention should not be so central that it is immune to all but the most dramatic of external changes. Although changes in more central processes such as values do occur, they occur incrementally over long periods of time (Roberts & DelVecchio, 2000) and may be quite hard to achieve as a result of a relatively brief laboratory procedure (Rokeach, 1973). The optimal process to target may thus be one that is moderately central to how people view themselves. One possible candidate is knowledge. People become invested in certain views of themselves and the world (Lord, Ross, & Lepper, 1979; Swann et al., 1987), and yet knowledge does change in response to evidence.

Evidence for the effectiveness of a knowledge-based intervention for producing long-term changes in psychological processes related to social disparities stems from work on the prejudice habit-breaking intervention (Devine, Forscher, Austin, & Cox, 2012). The habit-

breaking intervention is based on the premise that biases that occur unintentionally are like unwanted habits that can be broken through a combination of motivation to reduce unintentional bias, awareness of its existence and the ways it affects minorities, and effort practicing strategies that reduce unintentional bias (Devine, 1989). The intervention consists of a semi-interactive, self-paced slideshow composed of feedback, education, and a training section.

During the *feedback section*, the participants complete a task that can reveal how bias can operate unintentionally, like a race IAT, and receive feedback about their performance. The implicit measure is thus used to provide participants with a palpable experience of what unintentional bias feels like (Monteith, Voils, & Ashburn-Nardo, 2001) and a pedagogical tool to make people aware that not all of their responses are consistent with their intentions. Thus, although an implicit measure is an important component of the habit-breaking intervention, its importance stems from its pedagogical usefulness rather than because it is the habit-breaking intervention's primary target of change.

During the *education section*, the participants learn how some of our reactions are inconsistent with intentions. The slideshow applies this general principle to idea to social groups by showing how subtle stereotypic messages pervade the social environment, and how these messages can spontaneously pop to mind. The spontaneous reactions can then influence behavior unintentionally. This section then provides evidence that, although they can be subtle, these unintentionally biased behaviors they can still lead to adverse outcomes for minorities in a wide variety of domains. The education section is thus designed to leverage the momentary awareness provoked by completing the implicit measure into a more general, enduring awareness of how unintentional biases systematically disadvantage minorities. Following the education section, the participants complete a *training section* that describes evidence-based strategies, such as stereotype replacement (Monteith, 1993), individuating (Brewer, 1988; Fiske & Neuberg, 1990), perspective-taking (Galinsky & Moskowitz, 2000; Todd et al., 2011), and intergroup contact (Pettigrew, 1998; Pettigrew & Tropp, 2006), that, if practiced, can lead to reductions in unintentional bias.

In our original demonstration of the habit-breaking intervention's effects, people who received the habit-breaking intervention exhibited reductions in measured implicit bias that endured two months after receiving the intervention. In subsequent replications of this study, people who received the habit-breaking intervention were no different in their measured implicit bias than control participants, suggesting that our original results may have been a false positive (Forscher, Mitamura, et al., under review; Forscher, Cox, et al., in preparation).

Nevertheless, we have accumulated substantial evidence that the habit-breaking intervention does affect participants' concern about discrimination (Devine et al., 2012; Forscher, Mitamura, et al., under review; Forscher, Cox, et al., in preparation). The changes in concern are themselves associated with an increased tendency to notice when others act with bias and to biased behaviors in themselves and others as wrong (Forscher, Mitamura, et al., under review). These results suggest that the habit-breaking intervention changes people's knowledge about unintentional bias and its consequences for minorities. These changes lead to changes in how people perceive incidents in the social environment, which in turn can lead to further changes in knowledge. The initial changes in knowledge may thus create a recursive, self-sustaining feedback loop, which may be critical for sustaining any changes that the intervention creates (Forscher, Mitamura, et al., under review). This feedback loop may explain why the habit-breaking intervention's effects on knowledge endure and may be critical for creating

enduring changes in behavior. Indeed, the habit-breaking's effects on behavior may be especially enduring; in one sample, people who received the habit-breaking intervention were more likely to post a comment objecting to an online essay arguing that stereotyping is harmless than were control participants (Forscher, Mitamura, et al., under review). This difference occurred two years after the intervention was administered.

We have also gathered some evidence that knowledge-based habit-breaking may be effective at reducing some kinds of intergroup disparities. This evidence comes from a cluster-randomized trial in 92 academic departments of a 2.5 hour workshop based on the prejudice habit-breaking intervention (Carnes et al., 2015). The workshop was designed to improve climate for female faculty and thereby improve the representation of women in these departments. Three months after the workshop, faculty in experimental departments reported feeling more self-efficacy to promote gender equity than faculty in control departments (Carnes et al., 2015). Three years after the workshop, 52% of the new hires in experimental departments were women, compared to 28% in control departments (Devine, Forscher, Cox, Sheridan, Kaatz, & Carnes, in preparation). In addition, there is evidence that these differences in hiring rates were due to the influence of faculty who attended the habit-breaking workshop; faculty who worked with workshop attenders on department-related affairs (e.g., hiring committees) reported doing more to promote gender equity in their departments (Forscher et al., in preparation).

In sum, because knowledge is moderately related to the self-concept, interventions that change knowledge may be especially useful for reducing intergroup disparities. Knowledge requires less effort to change than highly central processes, such as values, but is more likely to sustain long-term change than are peripheral processes, such as implicit bias. Indeed, we believe our intervention is effective because it helps people to recognize that wanting to be free of bias is not equivalent to being free of bias. Knowledge that they could be, however unwittingly, complicit in the perpetuation of bias is critical to creating concern about discrimination. And that concern appears to change their understanding of discrimination and encourages them to see bias in themselves and in the world around them (e.g., in others, on TV, etc.). Although we have not directly tested the utility of knowledge-based interventions for changing disparities in legal outcomes, we believe these interventions are more likely to prove useful than are interventions that target implicit bias directly.

Resolving social disparities is not easy. We have argued that addressing issues of such societal importance requires the engagement of the conscious mind (Devine, 1989; Devine et al., 2012). These problems cannot be reduced to responses on any one measure of implicit bias, and assuming that they can respects neither their complexity nor seriousness.

Conclusion

We have argued that the current research on implicit bias does not speak strongly to whether implicit bias causes disparities in legal outcomes, nor does it speak directly to whether implicit bias is a useful target for intervention. In fact, interventions that attempt to directly change responses on measures of implicit bias may be doomed to fail because implicit bias is, by definition, peripheral to the self-concept. This means that, even if an implicit bias intervention reliably produces a change in implicit bias, this change is likely to fade due to the very forces in the social environment that create and sustain the biases in the first place. We believe that interventions that target processes more central to the self-concept, such as the knowledge-based

prejudice habit-breaking intervention, are more likely to produce the type of changes necessary to reduce legal disparities.

To be clear, although we are skeptical as to whether implicit bias plays a central role in creating and sustaining disparities in law, we do not mean to suggest that no forms of unintentional bias contribute to disparities in law. As we noted earlier in the chapter, unintentional biases occur for many reasons besides implicit bias – for example, established laws and procedures that advantage one group at the expense of another, unequal distributions of wealth and education, and/or a lack of awareness of the systematic advantages provided to some groups and not others. However, because each of these potential causes has different implications for which interventions are likely to be effective at resolving social disparities, it is important to distinguish between these potential causes rather than lumping them into the same category.

Finally, although the habit-breaking intervention has proven useful at producing enduring changes in some contexts, it has not yet been tested in a legal context. Indeed, it is quite possible that disparities in legal outcomes are better addressed through changes in policy than through individually-targeted psychological interventions. However, we will only discover the approaches that are effective through research that compares the relative effectiveness of different approaches. Considering the social importance of addressing disparities in legal outcomes, conducting this research is a high priority for future work.

References

Brewer, M. (1988). A dual-process model of impression formation. In T. Srull & R. Wyer (Eds.), *Advances in social cognition* (pp. 1–36). Hillsdale N.J.: L. Erlbaum Associates.

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (don't expect an easy answer). *Journal of Personality and Social Psychology*, *98*, 550–558.

Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential Priming Measures of Implicit Social Cognition: A Meta-Analysis of Associations With Behavior and Explicit Attitudes. *Personality and Social Psychology Review*, *16*, 330–350.

Carnes, M., Devine, P. G., Baier Manwell, L., Byars-Winston, A., Fine, E., Ford, C. E., … Sheridan, J. (2015). The effect of an intervention to break the gender bias habit for faculty at one institution: A cluster randomized, controlled trial. *Academic Medicine*, *90*, 221–230.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18.

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*, *48*, 1267–1278.

Fazio, R. H., & Olson, M. A. (2014). The MODE model: Attitude-behavior processes as a function of motivation and opportunity. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Duel process theories of the social mind* (pp. 155–171). New York, NY: Guilford Press.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 23, pp. 1–74). Elsevier.

Forscher, P. S., Cox, W. T. L., Mitamura, C., Allison, C., Dix, E. L., & Devine, P. G. (*In preparation*). The robustness of the habit-breaking intervention.

Forscher, P. S., & Devine, P. G. (2015). Controlling the Influence of Stereotypes on One's Thoughts. In R. A. Scott & S. M. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Sciences* (pp. 1–12). Hoboken, NJ, USA: John Wiley & Sons, Inc.

Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (Under review). A meta-analysis of changes in implicit bias.

Forscher, P. S., Mitamura, C., Dix, E. L., Cox, W. T. L., & Devine, P. G. (Under review). Breaking the prejudice habit: Mechanisms, timecourse, and longevity.

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology*, *78*, 708–724.

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology*, *108*, 553–561.

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*, 1464–1480.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17–41.

Gross, S. R., & Mauro, R. (1989). *Death & discrimination: racial disparities in capital sentencing*. Boston: Northeastern University Press.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83.

Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist*, *52*, 1280–1300.

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.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*, 2098–2109.

Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology*, *65*, 469–485.

Monteith, M. J., Voils, C. I., & Ashburn-Nardo, L. (2001). Taking a look underground: Detecting, interpreting, and reacting to implicit racial biases. *Social Cognition*, *19*, 395–417.

Mustard, D. B. (2001). Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts. *The Journal of Law and Economics*, *44*, 285–314.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*, 171–192.

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.

Pettigrew, T. F. (1998). Integroup contact theory. *Annual Review of Psychology*, *49*, 65–85.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*, 751–783.

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*, 968–973.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, *126*, 3–25.

Rokeach, M. (1973). *The nature of human values*. New York: The Free Press.

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, *89*, 845–851.

Swann, W. B., Griffin, J. J., Predmore, S. C., & Gaines, B. (1987). The cognitive-affective crossfire: When self-consistency confronts self-enhancement. *Journal of Personality and Social Psychology*, *52*, 881–889.

Tan, D. L., Morris, L., & Romero, J. (1996). Changes in attitude after diversity training. *Training & Development*, *50*, 54–56.

Todd, A. R., Bodenhausen, G. V., Richeson, J. A., & Galinsky, A. D. (2011). Perspective taking combats automatic expressions of racial bias. *Journal of Personality and Social Psychology*, *100*, 1027–1042.