

The prejudice habit-breaking intervention: An empowerment-based confrontation approach

12

William T.L. Cox and Patricia G. Devine

Department of Psychology, University of Wisconsin—Madison, Madison, WI, United States

Chapter Outline

Individual-level confrontations	250
Group-based training confrontations	251
Empowerment-based confrontation	254
Creating awareness and concern in sincere egalitarians	255
Channeling effort in effective directions	257
Evidence of changes within individuals	261
Evidence of individuals changing social contexts	263
Applications	265
Acknowledgments	268
References	268

On April 12, 2018, two Black men were seated at a table at Starbucks, waiting for a friend to join them before they ordered. A manager at the coffee shop decided this situation warranted police involvement and called the police who arrested the two Black men (Hauser, 2018). Labeled by many as a clear instance of racial bias, this event received abundant media attention and public discussion. This event and others like it are representative of systematic, historical patterns in which stereotyped group members have been treated with disrespect, intimidation, and even violence (see also Broman, Mavaddat, & Hsu, 2000; Dovidio, Hewstone, Glick, & Esses, 2010; Lebron, 2018; Nicas, 2018; Perry et al., 2016; Wakabayashi, 2017). Increased public attention to these events spurred discussion about both the ill effects of discrimination and how to most effectively confront and address bias.

Public attention to bias and the increasing encouragement to root out and confront bias encourages a careful look at how to best create opportunities for bias confrontation to be successful. Confrontations can take many forms. Sometimes confrontations involve individuals speaking up and challenging a specific instance of bias in another person (see Czopp, 2019; Monteith et al., 2019). Other times confrontations involve institutions offering (or requiring) diversity or bias reduction

training that confront trainees with ways in which the organization may be falling short of its diversity and inclusion goals (see [Pendry, Driscoll, & Field, 2007](#)). These individual-level and group-based training approaches are the most prevalent confrontation approaches, but in recent years, a third category has begun to emerge that is related to but distinct from these approaches — empowerment-based confrontation.

In this chapter, we briefly review insights from the individual-level confrontation work and the group-based trainings regarding both the factors that enhance and the factors that interfere with successful confrontations. We then discuss the third type of approach, empowerment-based confrontation, in greater detail. Empowerment-based confrontation approaches have begun to show encouraging evidence in recent years (e.g., [Carnes et al., 2015](#); [Devine, Forscher, Austin, & Cox, 2012](#); [Devine et al., 2017](#); see also [Hennes et al., 2018](#); [Moss-Racusin et al., 2018](#); [Pietri et al., 2017](#)). Our work in this area has been focused on the prejudice habit model ([Devine, 1989](#)), which led to the development and testing of the prejudice habit-breaking intervention ([Devine et al., 2012](#)). This operationalization of empowerment-based confrontation has been successful in helping individuals to confront and take steps to reduce bias within themselves and to increase inclusion within their social contexts.

Individual-level confrontations

As a number of chapters in this book attest, the impact of individual level, person-to-person, confrontations has received considerable empirical attention in recent years (see in particular [Czopp, 2019](#); [Monteith, Burns, & Hillbrand, 2019](#), for excellent reviews). Czopp, for example, reviews extant research on the intra- and interpersonal consequences of confrontation for the experiences of those who confront another about bias and those who are confronted about their expression of bias. Czopp also reviews research examining the impact of confrontation on those who, though not directly involved in the immediate confrontation, observe the confrontation episode. Monteith et al. focus more specifically on the challenge of would-be confronters who may wish to call out incidents of bias but may experience uncertainties about exactly what to say or what to do to confront in a way that will be successful in decreasing future biased behavior and will allow the confronters to be viewed favorably. In addressing these issues, Monteith et al.'s goal is to offer guidance on how to successfully navigate these challenges of individual-level confrontation situations.

Although a thorough review of the individual-level confrontation literature is beyond the scope of the present chapter, it is instructive to highlight a few important themes that emerge therein. One clear theme that emerges in the confrontation literature is that some confronters are more effective than others, such that confrontations are received more positively when they come from members of the nontargeted social group (e.g., White people and men) than from members of the group targeted by bias (e.g., Black people and women) (see [Czopp, 2019](#)).

Whereas target group members are often viewed as whiners and overly sensitive, the nontargeted group members are seen as having more credibility by virtue of not appearing to have an immediate vested interest in the outcome of the confrontation.

A second clear theme in individual-level confrontation literature is that confrontation settings create trepidations for would-be confronters. Staying silent, for example, effectively makes a would-be confronter complicit in allowing bias to occur unchecked, which can have negative personal consequences for the would-be confronter (failure to live up to one's values; e.g., [Shelton, Richeson, Salvatore, & Hill, 2006](#)) and the target of bias (e.g., directly experiencing the ill effects of the unchecked bias). However, actively confronting another's bias leaves the confronter vulnerable to negative reactions by the person confronted and any observers to the confrontation episode (see [Czopp, 2019](#); [Monteith et al., 2019](#)). A related concern is that people who are confronted about bias by others can become defensive and avoidant, which could disrupt the effectiveness of the confrontation.

Addressing defensiveness and avoidance highlights a third theme in the individual-level confrontation literature, namely that the style of confrontation is important and ultimately affects how successful confrontations are in reducing bias. Several studies have now shown that hostile, extreme, or threatening confrontations, though sometimes successful in reducing immediate expressions of bias, also lead to agitation and anger at others (e.g., [Czopp, Monteith, & Mark, 2006](#); [Martinez, Hebl, Smith, & Sabat, 2017](#); [Parker, Monteith, Moss-Racusin, & Van Camp, 2018](#)). Confrontations that are instead presented calmly and appeal to principles of fairness tend to mitigate the negative reactions on the part of those confronted ([Burns & Monteith, in press](#)). Finally, confrontations that are clear and tethered to specific evidence of bias are more effective in curbing expressions of bias than confrontations that are general or ambiguous regarding what was objectionable in the behavior of the person being confronted ([Parker et al., 2018](#)). Such specific and evidence-based confrontations lead confronters to experience heightened negative self-directed affect and to regulate their behavior to reduce bias in a subsequent setting where bias was possible (see also [Chaney & Sanchez, 2018](#)). [Monteith et al. \(2019\)](#) provide a number of suggestions to guide would-be confronters in creating confrontations.

Group-based training confrontations

Group-based diversity, cultural awareness, and implicit bias trainings have become big business in the United States as organizations have become both increasingly diverse and concerned with issues of climate and inclusion. The hope is that providing employees with training will enable them to effectively confront and address their own and others' biases within the organizations ([Dobbin & Kalev, 2013, 2016](#); [Paluck & Green, 2009](#); [Pendry et al., 2007](#)). Indeed, such trainings are very

often promoted in the aftermath of some critical incident that brings bias within the organization into focus, as was the case in the Starbucks incident discussed earlier, or after an analysis or report reveals that the organization is falling short in its diversity goals (Dobbin & Kalev, 2013, 2016). These trainings are often designed specifically to address a problem of bias in the organization with the assumption that individuals within the organization are both responsible for bias and have a responsibility to address the bias.

In a comprehensive review of diversity training approaches, Pendry et al. (2007) note that these trainings fall into four major categories: (1) providing trainees with information or enlightenment about social inequality, (2) eliciting emotional reactions by revealing that some members of the organization receive unfair advantages (e.g., White privilege), (3) emphasizing the trainees' common social identity (i.e., members of the organization) as opposed to the distinct social identities of trainees (i.e., different race, ethnic, or other groups) as a way to encourage a sense of "w-ness," or (4) providing trainees with immediate feedback regarding their own bias (e.g., using a task like the Implicit Association Test; IAT; Greenwald, Nosek, & Banaji, 2003), in the hopes that recognition of personal bias will make people aware of their biases and lead to efforts to reduce bias.

Throughout their review, Pendry et al. (2007) identify several challenges that create obstacles to diversity training having its intended effects. For example, participation in such trainings is often mandatory and the trainings are presented by "experts" from outside the organization who are meant to educate the trainees about the problems and encourage change. This push from outsiders is often experienced as threatening, eliciting defensiveness and anger. Pendry et al. (2007) note that these circumstances are especially problematic when diversity trainings are experienced as a confrontation that implies trainees are morally suspect, causing the trainings to backfire and to lead to backlash (see also Dobbin & Kalev, 2016; Pettigrew & Tropp, 2006, as cited in Paluck, 2006). Pendry et al. suggest that to mitigate these negative circumstances, trainers need to be both skilled and nimble in handling conflict. Trainers should be knowledgeable about the relevant psychological literature to explain processes underlying the ordinary nature of bias (Devine, 1989) and the possibility that bias could be overcome with effort (Devine, 1989; Monteith, 1993). Although some trainers may be skilled in conflict resolution and knowledgeable about the psychological literature, to the extent that they are not, their credibility is undermined and trainees may not take the training opportunity seriously.

A shared goal of most diversity trainings is to raise trainees' awareness of their vulnerability to displaying intergroup biases. The method for raising this type of awareness varies, but Pendry et al. caution that some of the exercises used to raise awareness are difficult to implement in a group setting and that trainees may not take the exercises seriously. For example, when taking the IAT, many trainees deny the validity of the feedback or reject it as a good measure of one's bias (Howell, Gaither, & Ratliff, 2015; Howell & Ratliff, 2017; Monteith, Voils, & Ashburn-Nardo, 2001). Alternatively, some tasks are more upsetting than the anticipated gains of motivating people to reduce bias (consider Elliott's famous

“blue eyes vs brown eyes” school activity; Elliott, 2003). Though Pendry et al. (2007) conclude that such experiences can be motivating for many trainees to work to overcome their biases, their review also reveals that diversity trainers have little to offer in the way of practical suggestions for how individuals can reduce their biases. This limitation is critically important precisely because the exercises can be tremendously effective at convincing people that their biases can leave them vulnerable to discriminating against others, however unwittingly. Creating awareness without providing any advice on how to reduce bias or create an inclusive environment leaves trainees ill-equipped to make progress on these fronts. Indeed, absent tools to help trainees reduce bias, trainees may feel helpless, because they are acutely aware of their vulnerability to express unintentional bias but do not know what to do differently. As a result, rather than improving climate and inclusion within organizations, training may result in a counter-intentional effect: the trainees may become avoidant of intergroup settings out of concern that they may show bias or do the wrong thing (Plant & Devine, 2003; Stephan & Stephan, 1985).

There are two additional striking and concerning features of extant diversity training that deserve attention. First, the selection of the diversity training strategy employed by any given diversity trainer appears to be largely a matter of personal preference or intuition about what is likely to work. The specific strategies reviewed in Pendry et al. (2007) are not grounded in a particular (or specific) theoretical analysis of what is required to make such trainings successful (though Pendry et al. made a compelling case that a theoretical analysis could be retrofitted to some of the approaches). Further, this state of affairs exists, at least in part, because unlike in the individual-level confrontation literature in which the negative and beneficial aspects of confrontations have been systematically investigated in experimental studies, there have been few attempts to evaluate the effectiveness of existing diversity training programs (Dobbin & Kalev, 2016; Pendry et al., 2007; Stephan & Stephan, 2001).

Though recognizing that such programs are well-intentioned, scholars have increasingly questioned whether diversity or bias trainings have gotten too far ahead of the evidence to suggest that they are helpful (Moss-Racusin et al., 2014), and some have explicitly challenged the ethics of implementing such trainings without evidence of their efficacy (Dobbin & Kalev, 2013, 2016; Nordell, 2017; Paluck, 2012). In a review of over 700 private-sector organizations, Kalev, Dobbin, and Kelly (2006), for example, suggested that diversity training programs can actually lead to less rather than more diversity within the organization — an outcome clearly at odds with the intentions of diversity training. Indeed, Paluck (2012) argues the current standards regarding the implementation of diversity trainings are simply not good enough and recommends that providing evidence that diversity or bias trainings work as intended “should be considered an ethical imperative, on the level of rigorous testing of medical interventions.”

In contrast to the individual-level confrontations, group-based diversity, cultural awareness, or bias trainings are typically not built on a strongly evidence-based model of change, and there have been few attempts to rigorously evaluate the

impact of such trainings. As such, the approaches are often rooted in wishful thinking and intuition, and little is known about their efficacy.

Empowerment-based confrontation

The remainder of this chapter focuses on empowerment-based confrontation. The goal of empowerment-based confrontation is to put people on a path to recognizing, confronting, and regulating bias within themselves. In contrast to individual-level confrontation approaches, empowerment-based approaches do not involve an immediate confrontation from another person in response to an instance of bias in the moment. Instead, this approach sets people up to confront bias in themselves in an effort to reduce their future expressions of bias. Similar to many group-based confrontation approaches, empowerment-based approaches increase people's awareness of their vulnerability to unintentional bias, but in contrast to many group-based trainings, empowerment-based approaches raise awareness in ways that mitigate defensiveness and also provide guidance, in the form of specific strategies to reduce bias.

One prominent example of empowerment-based confrontation is the prejudice habit-breaking intervention (Carnes et al., 2015; Devine et al., 2012), which is grounded in a strong theoretical analysis and a well-supported model of change (Amodio, Devine, & Harmon-Jones, 2007; Cox, Abramson, Devine, & Hollon, 2012; Devine, 1989; Devine et al., 1991; Monteith, 1993; Plant & Devine, 1998). Our review focuses on the prejudice habit-breaking intervention, which we have developed and tested over the past 10 years. The prejudice habit-breaking intervention has been experimentally assessed in several randomized-controlled studies and has been shown to create lasting, meaningful change among those whose conscious, earnestly held values oppose bias. We will review the key components of this intervention approach, and then review some of the empirical evidence showing its effectiveness (Devine, 1989; Devine et al., 2012; Prochaska & Velicer, 1997).

The prejudice habit-breaking intervention is a multifaceted educational presentation that has been implemented in two semiinteractive formats. The first is a computer presentation (Devine et al., 2012), and the second is an in-person seminar delivered by experts (Carnes et al., 2015). Across the formats, this intervention approach was built upon the adult learning literature (e.g., Prochaska & Velicer, 1997) and the prejudice habit model (Devine, 1989), which has been empirically assessed and supported in the research literature over the past 30 years (e.g., Amodio et al., 2007; Devine et al., 1991; Devine & Monteith, 1993; Plant & Devine 1998, 2009; Monteith, 1993; Monteith, Ashburn-Nardo, Voils, & Czopp, 2002). The prejudice habit model conceptualizes stereotypes and unintentional or implicit biases as "habits of mind" and frames overcoming bias as a process of "breaking the prejudice (or bias) habit" through conscious self-regulation. This approach emphasizes five key components: (1) sincere personal values that oppose bias, (2) awareness of one's vulnerability to unintentional biases, (3) concern about

the consequences of unintentional biases, (4) learning effective bias reduction strategies, and (5) self-sustained effort over time to reduce the influence of unintentional biases.

Creating awareness and concern in sincere egalitarians

Our approach focuses on people whose earnestly held personal values discourage prejudice and support egalitarianism (Devine, 1989). People with nonprejudiced values have the long-term goal of behaving in unbiased ways, and this goal is personally important to them (Devine, 1989; Monteith, 1993; Plant & Devine, 1998). These nonprejudiced values are a necessary first step for our approach, which relies upon individuals' sincere personal desire to behave in unbiased ways as the driving force for change. Those whose values support prejudice or are indifferent to prejudice and intergroup bias (Forscher, Cox, Graetz, & Devine, 2015) are, by definition, unmotivated to address or reduce bias.

Even people with strong personal values that oppose bias can fall prey to implicit or unintentional biases that arise from well-learned, automatically activated stereotypes (Devine, 1989). There is variability, however, in the extent to which people are aware of their own vulnerability to unintentional bias (Devine, 1989; Devine et al., 1991; Monteith et al., 2001; Plant & Devine, 2009). Without this awareness, there is no reason for someone to put effort into reducing bias. A crucial step, therefore, is confronting people with evidence that they may be vulnerable to expressing bias unintentionally, to the detriment of stereotyped group members.

Bias is inconsistent with an egalitarian's sense of self (Devine, 1989; Monteith, 1993), thus being confronted with evidence of unintentional bias should create dissonance that must be resolved. The prejudice habit-breaking intervention is designed to create and harness this dissonance to motivate the change process. There are three possible responses to this kind of confrontation. The first and, in our view, least likely response is that people will change their egalitarian values or the importance of those values to their sense of self (Devine, Tauer, Barron, Elliot, Vance, & Harmon-Jones, *in press*). This outcome is unlikely, both because people's core values are difficult to change and because if they do change, that would mean those were not personally important core values (and, those people, therefore, are not part of the intended audience for this approach). The second possible response is that the egalitarian person will reject the validity or veracity of the evidence that they may be vulnerable to bias (see also the review by Pendry et al., 2007). They may reject the notion that they are vulnerable to bias at all, or they may reject the idea that such biases are consequential. If they do not accept or believe that they could express consequential biases unintentionally, there will be no dissonance or perceived conflict with their egalitarian sense of self. Absent this dissonance, there is no motivation for people to try to reduce bias.

If, however, people are confronted in a way that leads them to accept that they may be vulnerable to consequential, unintentional biases, they will have the third possible response — they will be motivated to put effort into reducing the influence of unintentional biases. The prejudice habit-breaking intervention is crafted to guide

people to this third possible response, resulting in awareness and concern that motivate egalitarian people to put effort into reducing their unintentional bias. To reach this outcome, therefore, a key initial obstacle is confronting people in such a way that they do not reject the idea that they may be vulnerable to unintentional bias.

The intervention introduces people's personal vulnerability to unintentional bias in two ways. First, participants complete an implicit indicator of bias, the IAT (Greenwald et al., 2003) and receive feedback on their performance (e.g., based on their IAT scores, most non-Black people in the United States get feedback that they show a moderate or strong "automatic preference for White people over Black people"). In order to reduce the likelihood of participants rejecting the implications of this feedback, the intervention teaches them about the IAT as an indicator of unintentional bias and dispels common criticisms of the IAT (e.g., that its results are driven by block order effects or general color associations, rather than race).¹ Second, participants are presented with simple, relatable examples of mental habits that reveal a conflict between one's intended and unintended responses. The computerized intervention, for example, provides a brief, realistic anecdote in which someone makes a spontaneous stereotypic assumption that a Black man is homeless. This assumption is revealed to be incorrect and inspires the story's main character to reflect on the origin and unfairness of the assumption. Using these examples, the intervention illustrates the ways in which learned associations become automatically activated to create unintentional, habitual responses. In this way, the prejudice habit-breaking intervention echoes the aforementioned work showing that confrontations are more effective when they are tied to specific evidence (e.g., Parker et al., 2018) by using real IAT feedback, relatable anecdotal examples of unintentional bias, and reviewing research evidence about unintentional bias.

This approach engages motivation and minimizes defensiveness by conceptualizing bias as a habit of mind arising from socialization and learning experiences. Specifically, the intervention explains that children are repeatedly and frequently exposed to biases and stereotypes, which are omnipresent in culture (e.g., on television, in movies), to the point that biased responses become the default, automatic, and habitual reactions to members of stereotyped groups. In this way, the habit model teaches that unintentional bias is rather ordinary, that the participants are not alone in their vulnerability to unintentional bias, and that having unintentional bias does not reflect a moral failing. Emphasizing the ordinariness of unintentional bias helps avoid the defensiveness that often follows from bias feedback (as noted by Pendry et al., 2007; see also Howell et al., 2015).

After participants come to understand the pervasiveness and ordinariness of habitual biases, they learn about the ways in which these biases can be highly

¹We are aware of recent debates about the IAT's predictive validity for behavior. In this intervention work, we use the IAT as a learning tool. As participants complete the IAT, they are aware of the biases in their response times. This experience, and the subsequent feedback, helps participants to believe they may be vulnerable to bias, and this effect is useful in the present context whether or not the IAT predicts behavior well.

consequential for members of stereotyped groups (e.g., in medical settings, hiring settings). In the in-person versions of the intervention, we expand on different ways that biases might manifest (e.g., creating confirmation biases or self-fulfilling prophecies) (Darley & Gross, 1983; Rosenthal & Jacobson, 1968). Reviewing these various “bias constructs” helps participants to tune into and look out for the various and subtle ways bias may come about and fosters a sense of bias literacy, to give participants a language with which to discuss various forms of bias.

The prejudice habit-breaking intervention is designed to help people become aware and concerned that they may be vulnerable to unintentional biases and therefore may be contributing, however unwittingly, to negative consequences faced by members of stereotyped groups. This awareness and concern is key to engaging personal motivation to make efforts for change. Past evidence demonstrates, for example, that when nonprejudiced people are confronted with credible evidence that they have or may express bias, they are willing to put effort into tasks (e.g., reading race-related magazine articles, practice on a race-related cognitive task) they believe will help to regulate or otherwise reduce the expression of bias (Amodio et al., 2007; Monteith, 1993; Monteith et al., 2002; Plant & Devine, 2009). The increased interest in and effort on these tasks demonstrates that people are motivated and willing to put effort into reducing bias when confrontations credibly reveal an inconsistency between earnest egalitarian values and automatic responses. Having instilled this motivation, the prejudice habit-breaking intervention next helps people to have realistic expectations about the change process and empowers them to put forth effort in effective ways.

Channeling effort in effective directions

Once people with egalitarian personal values are aware of and concerned about their vulnerability to unintentional bias, they will be motivated to put effort into bringing their automatic responses in line with their intentions. Where they direct these efforts, however, varies. At times, people may not know where to direct their effort, leading to helplessness (Crocker, Luhtanen, Broadnax, & Blaine, 1999; Dovidio, Kawakami, & Gaertner, 2002). As noted earlier, bringing people to the point of awareness of unintentional bias without providing strategies to address bias can even lead people to avoid members of stereotyped groups to preclude any possibility of behaving in biased ways (Plant & Devine, 2003; Stephan & Stephan, 1985).

Sometimes, people will put effort into bias reduction strategies that are highly intuitive, but in fact have counter-intentional, counter-productive effects. For example, stereotype suppression (Macrae, Bodenhausen, Milne, & Jetten, 1994) is an intuitive strategy that backfires, resulting in *more* rather than *less* stereotype-based bias. Another intuitive but ineffective strategy that people may employ is group-blindness (e.g., colorblindness), in which they attempt to ignore group membership altogether (Apfelbaum, Norton, & Sommers, 2012; Plaut, Thomas, & Goren, 2009). This strategy, however, is not possible in most everyday situations, in which race, gender, age, and other group statuses are readily discernible. Further, to the extent

that people take pride in their social group memberships, ignoring that aspect of people's identities may be seen as rude or insulting (Apfelbaum, Sommers, & Norton, 2008; Plaut, Thomas, Hurd, & Romano, 2018; Richeson & Nussbaum, 2004). Lastly and importantly, group-blindness backfires — in one study of workplace climate, for example, the more White people in an organization ascribed to colorblind ideologies, the more biased they appeared to their Black coworkers, and the more their Black coworkers felt unwelcome in the organization (Plaut et al., 2009). One other intuitive but counter-productive strategy is having too strong a sense of one's own objectivity (Uhlmann & Cohen, 2007). Although objectivity seems antithetical to bias, personal confidence in one's objectivity leads to less questioning of one's judgments, often resulting in more, rather than less bias (Uhlmann & Cohen, 2007). To the extent that egalitarian, motivated people reach for these or other counter-productive strategies, their efforts at reducing bias are wasted.

The prejudice habit-breaking intervention, therefore, is designed to guide people away from ineffective strategies and teach people a toolkit of effective, evidence-based bias reduction strategies. Each of these strategies, drawn from the empirical literature, had previously been shown to reduce the expression of bias, at least in the short-term. Most often, these strategies (e.g., stereotype replacement, perspective-taking) were tested in isolation and were implemented at the behest of an experimenter (Blair, Ma, & Lenton, 2001; Brewer, 1988; Fiske & Neuberg, 1990; Galinsky & Moskowitz, 2000; Monteith, 1993; Pettigrew, 1998; Pettigrew & Tropp, 2006). In this way, the prior literature involved participants implementing a single strategy, because they were told to do so by an experimenter, rather than voluntarily using the strategies in the service of a long-term goal to reduce bias, and without participants being aware that the strategy may help reduce bias. In contrast, the prejudice habit-breaking intervention presents these strategies together as a toolkit that people can intentionally draw upon, if motivated, to help them progress toward their goal of reducing bias.

Each strategy (see Table 12.1) is described in straightforward, simple language, and then the computer presentation or human presenters discuss how to apply the strategy in different situations and share anecdotal or research-based examples of how the strategy can be useful at mitigating bias. Some strategies are well-suited to addressing bias in the moment after a biased thought has occurred (e.g., stereotype replacement), and some of the strategies are activities that can proactively reduce the likelihood of later expressions of bias (e.g., increasing intergroup contact). The strategies are meant to be broadly applicable to many situations, but no one strategy could fit all situations in which bias may be a concern. Overall, the strategies fit a general theme of helping people to slow down their thinking and rely less on “gut” reactions, which are often influenced by stereotypes and biases.

Another important feature of the toolkit of strategies is that the strategies are synergistic — if a White person uses the intergroup contact strategy and makes more friends who are Black people, that White person will have a better understanding of their friends' lives, and will be better equipped to engage in perspective-taking. Time spent practicing stereotype replacement should make the

Table 12.1 Strategies included in the prejudice habit-breaking intervention

Strategy	Brief description	Research adapted from	Status
Stereotype replacement	Detecting stereotypic thoughts or media portrayals and replacing them with nonstereotypic responses	Devine (1989), Devine and Monteith (1993), Monteith (1993)	Used in all iterations of the prejudice habit-breaking intervention
Counter-stereotypic imagery	Spending time thinking about counter-stereotypic exemplars	Blair et al. (2001), Dasgupta and Greenwald (2001)	Used in Devine et al. (2012), Carnes et al. (2015), and Forscher, Mitamura, Dix, Cox, and Devine (2017). Data in Forscher et al. (2017) indicated that this strategy may backfire, leading people to be less concerned about bias (we surmise it leads to thinking such as, “if Obama can be president, Black people must not have such a hard time!”). Subsequent iterations have excluded this strategy
Perspective-taking	Imagining what it would feel like to be in the position of a member of a different group	Galinsky and Moskowitz (2000)	Used in all iterations of the prejudice habit-breaking intervention
Individuation	Gather and focus on information that makes someone an individual, rather than generalizing based on group membership	Brewer (1988), Fiske and Neuberg (1990)	Used in all iterations of the prejudice habit-breaking intervention

(Continued)

Table 12.1 (Continued)

Strategy	Brief description	Research adapted from	Status
Increasing intergroup contact	Making positive, personal contact with members of stereotyped groups	Pettigrew (1998), Pettigrew and Tropp (2006)	Used in all iterations of the prejudice habit-breaking intervention
Consider situational explanations for behavior	Actively consider situational explanations for behavior, rather than jumping to dispositional explanations	Kawakami, Dovidio, Moll, Hermsen, and Russin (2000)	Briefly mentioned in the published in-person intervention work, and expanded to be a formal strategy to replace counter-stereotypic imagery in all current iterations
Modify your environment	Increase representation of underrepresented group members in your visual environment	Cheryan, Plaut, Davies, and Steele (2009)	Added to current iterations of the prejudice habit-breaking intervention
Committing to credentials	Commit to evaluation of criteria in advance of viewing any candidate	Uhlmann and Cohen (2005)	Added to current iterations of the prejudice habit-breaking intervention

The prejudice habit-breaking intervention teaches participants a set of bias-reducing strategies drawn from the empirical literature. As we test subsequent iterations of the prejudice habit-breaking intervention, we seek to expand the strategies. The strategies in the last three rows have been added in ongoing, currently unpublished tests of the intervention.

stereotype replacement process more efficient, which would make it easier to avoid generalizations and focus on individuating information. The intervention points out some of these synergistic possibilities, and encourages participants to think of different situations in which each strategy could be useful.

Because stereotypes and unintentional biases are deeply ingrained by the time people start trying to regulate their influence, reduction of unintentional bias is unlikely to happen quickly (Devine, 1989), even if people are armed with effective strategies. Instead, changing well-learned cognitive patterns requires sustained effort over time (Cox et al., 2012; Devine, 1989). If people, however, mistakenly believe that bias reduction should come quickly or easily, failure experiences in which they express bias may lead to helplessness or giving up (Dweck, 2006; Prochaska & Velicer, 1997; Rattan & Dweck, 2010). To encourage sustained effort and set realistic expectations, the prejudice habit-breaking intervention explicitly teaches people the prejudice habit model (Devine, 1989), in addition to elements of the model being infused throughout the presentation of the intervention. Participants are taught that, like breaking any habit, the process of breaking the prejudice habit does not occur instantaneously and requires sustained effort. Further, the intervention explicitly notes that people will likely “slip up” and express bias but assures them that if they return to the strategies, they will make progress and reduce bias. We also encourage participants, noting that, with practice, the strategies themselves can become more automatic, leading to habits of mind that oppose rather than promote bias. This understanding of the change process is meant to help encourage self-efficacy and resilience to future setbacks. In sum, our intervention approach (see Table 12.2) is multifaceted, with several theoretically motivated components that, together, are designed to empower lasting change by facilitating individuals to be open and able to confront and address bias in themselves, in a self-sustaining, enduring way.

Evidence of changes within individuals

We have assessed the effectiveness of the prejudice habit-breaking intervention across several randomized-controlled studies. Of key interest was whether the intervention caused changes in outcomes crucial to the expression and regulation of bias, specifically (1) awareness of personal vulnerability to unintentional bias, (2) concern that discrimination is a serious problem, and (3) self-efficacy to reduce bias within oneself. Across several studies, intervention participants increased in their reported awareness of their potential to express bias compared to control participants; this increased awareness was observed at 6 weeks (Devine et al., 2012), 3 months (Carnes et al., 2015), and up to 2 years (Forscher et al., 2017) after participants completed the intervention.

The intervention also led to increases in the extent to which participants were concerned about racial discrimination as a serious social problem that needs to be addressed (Devine et al., 2012), which also lasted up to 2 years (Forscher et al., 2017). Additionally, participants reported higher levels of self-efficacy to recognize and replace stereotypic responses (Carnes et al., 2015), and were more likely to

Table 12.2 Conceptual and operationalized components of the prejudice habit-breaking intervention

Conceptual purpose	Computerized intervention	In-person intervention	Learning objective
	(Devine et al., 2012; Forscher et al., 2017)	(Carnes et al., 2015; Devine et al., 2017)	
<p><i>Awareness</i> Introduce personal vulnerability to unintentional bias</p>	IAT feedback; anecdote about making a spontaneous stereotypic assumption	IAT feedback; examples of cognitive and perceptual habits/interference (e.g., Stroop task)	Understand that bias is ordinary, pervasive, and not a moral failing
<p><i>Concern</i> Reveal that unintentional bias is consequential</p>	Review research evidence showing intergroup disparities thought to arise from unintentional processes	Review research evidence about different bias constructs (e.g., prescriptive norms, reconstructing credentials) that contribute to intergroup disparities thought to arise from unintentional processes	
<p><i>The habit model</i> Understand the model of change</p>	Review the steps in breaking a habit: motivation, awareness, strategies, effort		Understand that bias is a habit that can be broken; have realistic expectations about the habit-breaking process (i.e., it requires effort over time and likely will involve setbacks)
<p><i>Strategies</i> Channel effort into productive strategies</p>	Teach a set of evidence-based strategies for reducing bias (see Table 12.1 for details)		

The theoretical components of the prejudice habit-breaking intervention, the ways in which they have been operationalized in published work, and the concomitant learning objectives. IAT, Implicit Association Test.

notice and label bias in themselves, in others, and in the media (Forscher et al., 2017). These outcomes provide compelling initial evidence that the prejudice habit-breaking intervention is effective at empowering people to create lasting changes within themselves.

We also explored the extent to which the intervention altered the expression of automatic biases. With regard to this issue, the evidence is somewhat mixed. In our initial test of the prejudice habit-breaking intervention (Devine et al., 2012) and a recent replication (Cox, Dix, Scott, & Devine, unpublished manuscript), intervention participants decreased in their level of implicit bias, as measured by the IAT, whereas control participants did not. In both of these studies, these effects lasted out to 8 weeks postintervention. In a third replication study, however, IAT scores decreased for both intervention and control participants (Forscher et al., 2017). This study was designed to test the time-course of the reduction in implicit bias, and involved frequent follow-up measures, with participants completing the race evaluative IAT every other day for 2 weeks (Forscher et al., 2017). It is possible that the control participants' IAT scores decreased as a result of a practice effect on the IAT (e.g., see Keatley, Clarke, Ferguson, & Hagger, 2014). Across all three studies, however, intervention participants' IAT scores decreased.

The precise interpretation of this pattern is complicated, even before recent debates about the IAT's predictive validity (Greenwald, Banaji, & Nosek, 2015; Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2015). Further, it is important to note that reduction in bias on the IAT may not translate to reduction in biased behavior (Burns, Monteith, & Parker, 2017; Forscher, Lai, et al., under revision). Regardless of whether the IAT is a reliable predictor of other behaviors, it remains a useful measure of automatic race-related responses, and there are several ways the prejudice habit-breaking intervention may have led to reductions in automatic race bias. It could be that the effort participants put into regulating their bias over time led to weakened negative automatic race associations or to the development or strengthening of positive race associations, either of which would lead to reduced IAT scores. Another possibility is that the automatic race associations remained relatively unchanged, but intervention participants' ability to exert control over automatic responses became more efficient or automatic (Amodio, Devine, & Harmon-Jones, 2008), resulting in better (less biased) performance on the IAT. Further replication and examination of potential mechanisms of this effect will be addressed in future research.

Evidence of individuals changing social contexts

Although focused on empowering individuals to become agents of change within themselves, some of our evidence indicates they also become agents of change in the social contexts in which they live and work. For example, intervention participants are more likely to self-report that creating a welcoming environment for members of all social groups is their personal responsibility (Cox et al., unpublished manuscript). We also sought to extend beyond self-report measures, to examine whether intervention participants were more likely to speak out against bias in the

social environment. Two years after either receiving the intervention or being assigned to the control group, we contacted participants for a long-term behavioral follow-up study. Participants were led to believe that they were part of a study conducted by our university's school newspaper and did not know that this study was related to the intervention study. They were told that the newspaper was testing out a new editorial format, in which one student writes an opinion piece, and other students respond to it. The editorial that participants read lauded stereotypes as harmless and useful, saying that stereotypes had gotten a "bad rap" in modern "PC culture" and made other arguments in favor of stereotypes and stereotyping. Participants rated their agreement/disagreement with the article and were given the opportunity to write a response to the editorial that would ostensibly be published in the school paper with the editorial, if they chose to write a response.

Importantly, both intervention and control participants disagreed with the content of the article to the same extent — most everyone disagreed with what the article had to say. But, intervention participants were more likely than control participants to translate that disagreement into action, voluntarily writing a response that confronted the author of the bias-promoting rhetoric. Often, these participants offered concrete criticism about how stereotypes cause harm for members of stereotyped groups (Forscher et al., 2017). Even though it had been 2 years since it was administered, the prejudice habit-breaking intervention had enduring effects on participants' behavior, making them more likely to confront bias in a public forum. These patterns provide initial evidence that the prejudice habit-breaking intervention inspires people to take action not only against potential bias within themselves, but in the world around them.

One specific context in which bias is thought to be a major issue is in science, technology, engineering, and math (STEM) fields. Many national and international organizations have identified bias and lack of diversity as key obstacles in STEM (Moss-Racusin et al., 2014; NSF, 2007). STEM is often seen as being unwelcoming to women and people of color, and, correspondingly, members of those groups are underrepresented in STEM professions (LaCosse, Sekaquaptewa, & Bennett, 2016; Moss-Racusin et al., 2014). We sought to extend our test of the prejudice habit-breaking intervention in the STEM context, focused specifically on gender bias (Carnes et al., 2015). In a large-scale, cluster-randomized trial, we randomly assigned STEM departments to receive a gender version of the habit-breaking intervention or to serve as controls (Carnes et al., 2015). We assessed potential effects on departmental climate by accessing data from the *Study of Faculty Worklife Survey*, a campus climate survey that is regularly administered to faculty and was unrelated to the intervention study. In this campus climate survey, both male and female faculty in intervention departments self-reported (1) feeling better fit in their departments, (2) that their work was more valued, and (3) higher levels of comfort about raising family obligations than faculty in control departments. We speculate that these climate improvements arose because the prejudice habit-breaking intervention provided the faculty with a common language and understanding of issues related to unintentional gender bias in STEM, which in turn led to more productive discussions around these issues. Further, we are encouraged by the evidence

showing that climate improved for male faculty as well. It may be the case that lack of self-efficacy to address or discuss gender bias issues causes tension for male and female faculty, and the intervention releases that tension, to improve the climate for everyone. These speculations are supported by anecdotal feedback from intervention participants and will be the subject of future systematic research on how the prejudice habit-breaking intervention's effects operate within organizations.

One follow-up study with the STEM departments suggested that faculty who did not attend the in-person intervention, but who work closely with a faculty member who did attend, self-reported taking more actions related to gender equity than faculty who did not work closely with an attendee (Forscher, Carnes, Sheridan, & Devine, unpublished manuscript). This finding seems to imply that the intervention's effects spread within organizations even to those who do not receive the intervention training.

Because a key institutional-level concern in STEM is the underrepresentation of women, we recently assessed the extent to which the prejudice habit-breaking intervention led to differences in hiring patterns (Devine et al., 2017). In the 2 years before our study began, the percentages of new faculty hires who were female were comparable in the intervention (32%) and control (33%) departments. In the 2 years following our study, the percentage of new female faculty hires did not change in control departments (32%), but increased by 15% points in intervention departments (47%). See Fig. 12.1. These patterns provide initial but promising evidence that the prejudice habit-breaking intervention, although focused on individual-level change, also has beneficial effects at the institutional level.

The effects of the intervention on academic STEM department climate and increased representation of women are especially exciting and important, but, to date, we have not fully explored how the intervention produces these changes. It is possible that hiring committees were more (1) aware of, (2) concerned about, and (3) equipped with strategies to address unintentional bias in their review of applicants, resulting in more women being offered positions. Or, perhaps control and intervention departments offered positions to equal numbers of women, but women were more likely to accept the offers from intervention departments because they perceived the more positive climates in the intervention departments, noted above. Further work will seek to unpack these possible mechanisms. We should also note that being hired into a STEM department is only a first step for women faculty. For the intervention to be fully successful, women faculty must succeed at similar rates as their men colleagues (e.g., obtain promotions, be awarded grants, rise to leadership positions).

Applications

Although the prejudice habit-breaking intervention's success in improving some self-report and behavioral outcomes is promising thus far, subsequent research is

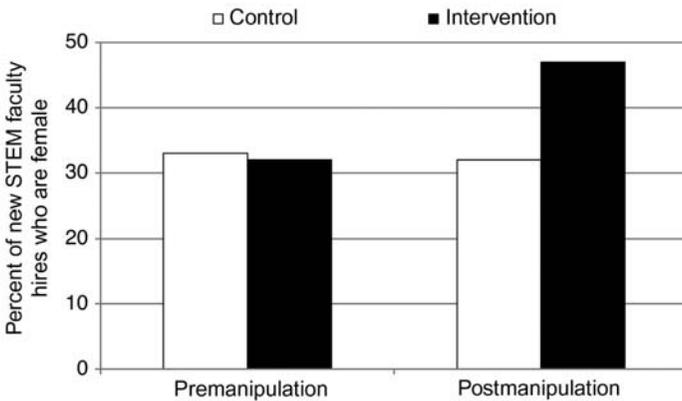


Figure 12.1 Effects of gender habit-breaking intervention on STEM faculty hires. The prejudice habit-breaking intervention was adapted for the gender/STEM context and tested in a cluster-randomized experiment using academic STEM departments. In the 2 years prior to the study, the proportion of new hires who were women was comparable for both intervention and control departments. After the study, control department rates remained unchanged, but rates in intervention departments increased (Devine et al., 2017). *STEM*, Science, technology, engineering, and math.

needed to broaden the set of outcomes explored. Because an empowerment-based approach is focused specifically on helping people to become self-confronters, we would expect to see effects on a wide range of outcomes related to bias, in a wide variety of contexts. One important extension is examining how intervention participants interact with members of stigmatized groups. Although our STEM study showed that the intervention created a more positive climate for men and women, none of our studies thus far have assessed participants' behavior during intergroup interactions, which is a common context in which bias plays out and has detrimental effects. We would also like to more deeply assess the on-line processes by which people recognize and regulate stereotyping or unintentional biases. Additional outcomes will enable both testing of the generalizability of the intervention's effects and assessment of the mechanisms by which it yields its effects, especially within organizations and social networks.

In addition to broadening the outcomes assessed, we and our colleagues have begun to broaden the contexts in which we evaluate the prejudice habit-breaking intervention. Specifically, we are exploring the impact of the intervention on grade school teachers (e.g., to reduce achievement gaps), police (e.g., to reduce disparities in traffic stops and use of force), doctors (e.g., to reduce disparities in race-based assumptions about patient diagnoses), and other groups. Because the prejudice habit-breaking intervention is focused on general mental processes, we argue that the fundamental principles should generalize to any context in which disparities are caused by automatic processes that influence behavior in opposition to intentions.

Further, it should be acknowledged that the prejudice habit-breaking intervention is multifaceted, with several interrelated components (e.g., awareness, concern, the

habit model, strategies). Each component is theoretically motivated, and according to our analysis, the full set of components is necessary for lasting change. We have not, however, directly tested the unique contribution(s) of the constituent components. Individual components in isolation could be sufficient for some people. For example, if someone is already aware of and concerned about their unintentional bias, perhaps learning the strategies and the habit model will be sufficient to empower their change process. These issues are being explored in ongoing work.

The prejudice habit-breaking intervention is only one instantiation of empowerment-based approaches. Another approach that we see fitting in this category is the recently developed “UNITE with Video Interventions for Diversity in STEM” program (UNITE with VIDS; [Hennes et al., 2018](#); see [Moss-Racusin et al., 2018](#)). Like the prejudice habit-breaking intervention, this approach raises awareness, teaches people that bias can be reduced, and provides tools to overcome bias. The focus of the UNITE with VIDS training is, so far as we are aware, solely on gender and STEM, but initial published evidence is promising. As this and other empowerment-based approaches are developed and rigorously tested, the field will be better equipped to understand the mechanisms by which empowerment-based confrontations can create lasting change, and researchers can continue to develop more effective approaches to reduce stereotyping, bias, and their behavioral manifestations.

Lastly and importantly, the prejudice habit-breaking intervention is catered specifically to people whose implicit bias belies their nonprejudiced intentions. To date, our samples have predominantly been comprised of people with egalitarian values. For people who lack those values but complete our training, we can imagine four possible outcomes. First, the intervention may serve to increase the salience of norms that prohibit prejudice, and nonegalitarian people may use the bias-reducing strategies in the service of conforming to those norms, thereby appearing or being less biased (e.g., similar to some individual-level confrontation studies ([Burns & Monteith, in press](#))). Second, it is possible that the intervention’s evidence-based, nonaggressive approach to discussing bias would engage nonegalitarian participants in such a way that their values would genuinely become more egalitarian. This process would start them on a path to change that is then similar to those who were egalitarian at the outset. Third, this approach may simply be ineffective for nonegalitarians, because automatic expressions of bias lead to outcomes that are consistent with their motivations and conscious values (e.g., [Forscher et al., 2015](#)). A fourth possibility is that, to the extent that the intervention is seen as imposing normative pressure to behave in less biased ways, nonegalitarians may resent the pressure and even manifest backlash ([Plant & Devine, 2001](#)). We have not yet explored these possibilities in our empirical work, but if the prejudice habit-breaking intervention leads to the latter two negative outcomes among nonegalitarians, effecting change in nonegalitarians may require a very different approach, perhaps one that changes their values or makes the expression of bias maladaptive for them (e.g., see [Tropp & Barlow, 2018](#)).

Considered from the lens of typical confrontation research in which confrontations are direct and primarily initiated by another person, our empowerment-based

confrontation approach is a bit unusual. Rather than being directly confronted by another person, people who experience the prejudice habit-breaking intervention enter a space in which they come to recognize vulnerability to unintentional bias in themselves, understand how it negatively affects others, and become empowered to address it. This empowerment-based confrontation approach therefore effectively makes the self the primary confronter over time.

Prejudice is a multilevel, complex problem, and no one approach is likely to be sufficient to address all forms of prejudice. Accordingly, the prejudice habit-breaking intervention is not meant to be a complete solution to reducing bias and disparities. Our approach does, however, show considerable promise for improving intergroup relations by helping well-intentioned people to bring their own behavior in line with their nonprejudiced personal values and to sustain that effort over time. By engaging egalitarian values and enlisting individuals as their own agents of change, we can extend the influence of a confrontation beyond the context in which it initially occurs, empowering people to carry forward what they learn from the confrontation to maintain a sustained process of self-confrontation, resulting in lasting bias reduction.

Acknowledgments

Preparation of this chapter was supported by an internal grant from the Office of the Provost of the University of Wisconsin–Madison awarded to both authors, as well as NIH grant 5R01GM111002-02 and a WARF Professorship awarded to the second author. We would also like to thank Laura I. Rivera for her help with the preparation of this manuscript.

References

- Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (2007). A dynamic model of guilt: Implications for motivation and self-regulation in the context of prejudice. *Psychological Science, 18*, 524–530. Available from <https://doi.org/10.1111/j.1467-9280.2007.01933.x>.
- Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (2008). Individual differences in the regulation of intergroup bias: The role of conflict monitoring and neural signals for control. *Journal of Personality and Social Psychology, 94*, 60–74. Available from <https://doi.org/10.1037/0022-3514.94.1.60>.
- Apfelbaum, E. P., Norton, M. I., & Sommers, S. R. (2012). Racial color blindness: Emergence, practice, and implications. *Current Directions in Psychological Science, 21*, 205–209. Available from <https://doi.org/10.1177/0963721411434980>.
- Apfelbaum, E. P., Sommers, S. R., & Norton, M. I. (2008). Seeing race and seeming racist? Evaluating strategic colorblindness in social interaction. *Journal of Personality and Social Psychology, 95*, 918–932. Available from <https://doi.org/10.1037/a0011990>.
- Blair, I. M., Ma, J., & Lenton, A. (2001). Imagining stereotypes away: The moderation of automatic stereotypes through mental imagery. *Journal of Personality and Social Psychology, 81*, 828–841. Available from <https://doi.org/10.1037/0022-3514.81.5.828>.

- Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull, & R. S. Wyer, Jr. (Eds.), *Advances in social cognition* (Vol. 1, pp. 1–36). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
- Broman, C. L., Mavaddat, R., & Hsu, S. Y. (2000). The experience and consequences of perceived racial discrimination: A study of African Americans. *Journal of Black Psychology*, 26, 165–180. Available from <https://doi.org/10.1007/s40615-015-0170-3>.
- Burns, M. D., Monteith, M. J., & Parker, L. R. (2017). Training away bias: The differential effects of counterstereotype training and self-regulation on stereotype activation and application. *Journal of Experimental Social Psychology*, 73, 97–110.
- Burns, M. D., & Monteith, M. J. (in press). Confronting stereotypic biases: Does internal versus external motivational framing matter? *Group Processes and Intergroup Relations*, 2018, 1368430218798041.
- Carnes, M., Devine, P. G., Manwell, L. B., Byars-Winston, A., Fine, E., Ford, C. E., . . . Sheridan, J. (2015). Effect of an intervention to break the gender bias habit: A cluster randomized, controlled trial. *Academic Medicine*, 90, 221–230. Available from <https://doi.org/10.1097/ACM.0000000000000552>.
- Chaney, K. E., & Sanchez, D. T. (2018). The endurance of interpersonal confrontations as a prejudice reduction strategy. *Personality and Social Psychology Bulletin*, 44, 418–429. Available from <https://doi.org/10.1177/0146167217741344>.
- Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology*, 97(6), 1045–1060.
- Cox, W. T. L., Abramson, L. Y., Devine, P. G., & Hollon, S. D. (2012). Stereotypes, prejudice, and depression: The integrated perspective. *Perspectives on Psychological Science*, 7, 427–449. Available from <https://doi.org/10.1177/1745691612455204>.
- Cox, W. T. L., Dix E. L., Scott K., & Devine P. G. (2018). *Campus-wide trials of a computerized intervention to break the prejudice habit*. Unpublished manuscript.
- Crocker, J., Luhtanen, R., Broadnax, S., & Blaine, B. E. (1999). Belief in U.S. government conspiracies against Blacks among Black and White college students: Powerlessness or system blame? *Personality and Social Psychology Bulletin*, 25, 941–953. Available from <https://doi.org/10.1177/01461672992511003>.
- Czopp, A. M. (2019). The consequences of confronting prejudice. In R. K. Mallett, & M. J. Monteith (Eds.), *Confronting prejudice and discrimination: The science of changing minds and behaviors*. Philadelphia, PA: Elsevier.
- Czopp, A. M., Monteith, M. J., & Mark, A. Y. (2006). Standing up for a change: Reducing bias through interpersonal confrontation. *Journal of Personality and Social Psychology*, 90(5), 784–803. Available from <https://doi.org/10.1037/0022-3514.90.5.784>.
- Darley, J. M., & Gross, P. G. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44, 20–33. Available from <https://doi.org/10.1037/0022-3514.81.5.800>.
- 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(5), 800–814. Available from <https://doi.org/10.1037/0022-3514.81.5.800>.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5–18. Available from <https://doi.org/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*, 48, 1267–1278. Available from <https://doi.org/10.1016/j.jesp.2012.06.003>.
- Devine, P. G., Forscher, P. S., Cox, W. T. L., Kaatz, A., Sheridan, J., & Carnes, M. (2017). A gender bias habit-breaking intervention led to increased hiring of female faculty in STEM departments. *Journal of Experimental Social Psychology*, 73, 211–215. Available from <https://doi.org/10.1016/j.jesp.2017.07.002>.
- Devine, P. G., Monteith, M. J., Zuwerink, J. R., & Elliot, A. J. (1991). Prejudice with and without compunction. *Journal of Personality and Social Psychology*, 60, 817–830.
- Devine, P. G., & Monteith, M. J. (1993). The role of discrepancy associated affect in prejudice reduction. In D. M. Mackie, & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping* (pp. 317–344). New York: Academic Press.
- Devine, P. G., Tauer, J. M., Barron, K. E., Elliot, A. J., Vance, K. M., & Harmon-Jones, E., (in press) Moving beyond attitude change in the study of dissonance-related processes: An update on the role of discomfort, In: Harmon-Jones, E., (Ed), *Cognitive dissonance: Re-examining a pivotal theory in psychology*, 2nd ed., American Psychological Association, Washington, DC.
- Dobbin, F., & Kalev, A. (2013). The origins and effects of corporate diversity programs. In Q. M. Roberson (Ed.), *The Oxford handbook of diversity and work* (pp. 253–281). Oxford, UK: Oxford University Press.
- Dobbin, F., & Kalev, A. (2016). Why diversity programs fail. *Harvard Business Review*, 94, 52–61.
- Dovidio, J. F., Hewstone, M., Glick, P., & Esses, V. M. (2010). Prejudice, stereotyping, and discrimination: Theoretical and empirical overview. In J. F. Dovidio, M. Hewstone, P. Glick, & V. M. Esses (Eds.), *SAGE handbook of prejudice, stereotyping, and discrimination* (pp. 3–28). London: Sage.
- Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, 82, 62–68. Available from <https://doi.org/10.1037/0022-3514.82.1.62>.
- Dweck, C. S. (2006). *Mindset*. New York, NY: Random House.
- Elliott, J. (2003). Yale University, WGBH (Television station: Boston, MA), & PBS DVD (Firm). *A class divided*. New Haven, CT: Yale University Films.
- 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* (23, pp. 1–74). San Diego, CA: Academic Press.
- Forscher, P. S. (Unpublished doctoral dissertation). <https://thesiscommons.org/4t7fy/>.
- Forscher, P. S., Cox, W. T. L., Graetz, N., & Devine, P. G. (2015). The motivation to express prejudice. *Journal of Personality and Social Psychology*, 109(5), 791–812. Available from <https://doi.org/10.1037/pspi0000030>.
- Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (under revision). *A meta-analysis of change in implicit bias*.
- Forscher, P. S., Mitamura, C., Dix, E. L., Cox, W. T. L., & Devine, P. G. (2017). Breaking the prejudice habit: Mechanisms, timecourse, and longevity. *Journal of Experimental Social Psychology*. Available from <https://doi.org/10.1016/j.jesp.2017.04.009>.
- 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(4), 708–724. Available from <https://doi.org/10.1037/0022-3514.78.4.708>.
- 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. Available from <https://doi.org/10.1037/pspa0000016>.

- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). 'Understanding and using the Implicit Association Test: I. an improved scoring algorithm': Correction to Greenwald et al. (2003). *Journal of Personality and Social Psychology*, 85, 481. Available from <https://doi.org/10.1037/h0087889>.
- 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. Available from <https://doi.org/10.1037/a0015575>.
- Hauser, C. (2018, April 19). *Men arrested at Starbucks hope to ensure 'This Situation Doesn't Happen Again'*. Retrieved from <https://www.nytimes.com/2018/04/19/us/starbucks-black-men-arrests-gma.html>.
- Hennes, E. P., Pietri, E. S., Moss-Racusin, C. A., Mason, K. A., Dovidio, J. F., Brescoll, A., ... Handelsman, J. (2018). Increasing the perceived malleability of gender bias using a modified video intervention for diversity in STEM (VIDS). *Group Processes & Intergroup Relations*, 21(5), 788–809. Available from <https://doi.org/10.1177/1368430218755923>.
- Howell, J. L., Gaither, S. E., & Ratliff, K. A. (2015). Caught in the middle: Defensive responses to IAT feedback among whites, blacks, and biracial black/whites. *Social Psychological and Personality Science*, 6(4), 373–381. Available from <https://doi.org/10.1177/1948550614561127>.
- Howell, J. L., & Ratliff, K. A. (2017). 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. Available from <https://doi.org/10.1111/bjso.12168>.
- Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *American Sociological Review*, 71, 589–617. Available from <https://doi.org/10.1177/000312240607100404>.
- Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotypic associations on stereotypic activation. *Journal of Personality and Social Psychology*, 78, 871–888. Available from <https://doi.org/10.1037/0022-3514.78.5.871>.
- Keatley, D. A., Clarke, D. D., Ferguson, E., & Hagger, M. S. (2014). Effects of pretesting implicit self-determined motivation on behavioural engagement: Evidence for the mere measurement effect at the implicit level. *Frontiers in Psychology*, 5(125). Available from <https://doi.org/10.3389/fpsyg.2014.00125>.
- LaCosse, J., Sekaquaptewa, D., & Bennett, J. (2016). STEM stereotypic attribution bias among women in an unwelcoming science setting. *Psychology of Women Quarterly*, 1, 20. Available from <https://doi.org/10.1177/0361684316630965>.
- Lebron, C. (February 05, 2018). *Who first showed us that black lives matter?* Retrieved from <https://www.nytimes.com/2018/02/05/opinion/black-lives-matter-philosophy.html>.
- Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of Personality and Social Psychology*, 67(5), 808–817. Available from <https://doi.org/10.1037/0022-3514.67.5.808>.
- Martinez, L., Hebl, M., Smith, N., & Sabat, I. (2017). Standing up and speaking out against prejudice toward gay men in the workplace. *Journal of Vocational Behavior*, 103, 71–85. Available from <https://doi.org/10.1016/j.jvb.2017.08.001>.
- 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. Available from <https://doi.org/10.1037/0022-3514.65.3.469>.
- Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of*

- Personality and Social Psychology*, 83, 1029–1050. Available from <https://doi.org/10.1037/0022-3514.83.5.1029>.
- Monteith, M. J., Burns, M. D., & Hildebrand, L. L. (2019). Navigating successful confrontations: What should I say and how should I say it? In R. K. Mallett, & M. J. Monteith (Eds.), *Confronting prejudice and discrimination: The science of changing minds and behaviors*. Philadelphia, PA: Elsevier.
- 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. Available from <https://doi.org/10.1521/soco.19.4.395.20759>.
- Moss-Racusin, C. A., Pietri, E. S., Hennes, E. P., Dovidio, J. F., Brescoll, V. L., Roussos, G., & Handelsman, J. (2018). Reducing STEM gender bias with VIDS (video interventions for diversity in STEM). *Journal of Experimental Psychology: Applied*, 24(2), 236–260. Available from <https://doi.org/10.1037/xap0000144>.
- Moss-Racusin, C. A., van der Toorn, J., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2014). Scientific diversity interventions. *Science*, 343, 615–616. Available from <https://doi.org/10.1126/science.1245936>.
- National Science Foundation. (2007). Women, minorities, and persons with disabilities in science and engineering. In *NSF 07–315*.
- Nicas, J. (2018, July 16). *Uber is target of federal sex discrimination inquiry*. Retrieved from <https://www.nytimes.com/2018/07/16/technology/uber-sex-discrimination-inquiry.html>.
- Nordell, J. (2017, May 07). *Is this how discrimination ends?* Retrieved from <https://www.theatlantic.com/science/archive/2017/05/unconscious-bias-training/525405/>.
- Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2015). Using the IAT to predict ethnic and racial discrimination: Small effect sizes of unknown societal significance. *Journal of Personality and Social Psychology*, 108, 562–571. Available from <https://doi.org/10.1037/pspa0000023>.
- Paluck, E. L. (2006). Diversity training and intergroup contact: A call to action research. *Journal of Social Issues*, 623, 439–451. Available from <https://doi.org/10.1111/j.1540-4560.2006.00474.x>.
- Paluck, E. L. (2012). Prejudice and conflict reduction interventions. In L. Tropp (Ed.), *Oxford handbook of intergroup conflict*. Oxford University Press.
- 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. Available from <https://doi.org/10.1111/j.1540-4560.2006.00474.x>.
- Parker, L. R., Monteith, M. J., Moss-Racusin, C. A., & Van Camp, A. R. (2018). Promoting concern about gender bias with evidence-based confrontation. *Journal of Experimental Social Psychology*, 74, 8–23. Available from <https://doi.org/10.1016/j.jesp.2017.07.009>.
- Pendry, L. F., Driscoll, D. M., & Field, S. C. (2007). Diversity training: Putting theory into practice. *Journal of Occupational and Organizational Psychology*, 80, 27–50. Available from <https://doi.org/10.1348/096317906X118397>.
- Perry, S. P., Hardeman, R., Burke, S. E., Cunningham, B., Burgess, D. J., & van Ryn, M. (2016). The impact of everyday discrimination and racial identity centrality on African American medical student well-being: A report from the Medical Student CHANGES study. *Journal of Racial and Ethnic Health Disparities*, 3(3), 519–526. Available from <https://doi.org/10.1007/s40615-015-0170-3>.
- Pettigrew, T. F. (1998). Applying social psychology to international social issues. *Journal of Social Issues*, 54(4), 663–675. Available from <https://doi.org/10.1111/j.1540-4560.1998.tb01242.x>.

- Pettigrew, T., & Tropp, L. (2006). Interpersonal relations and group processes. *Journal of Personality and Social Psychology*, *90*(5), 751–783. Available from <https://doi.org/10.1037/0022-3514.90.5.751>.
- Pietri, E. S., Moss-Racusin, C. A., Dovidio, J. F., Guha, D., Roussos, G., Brescoll, V. L., & Handelsman, J. (2017). Using video to increase gender bias literacy toward women in science. *Psychology of Women Quarterly*, *41*, 175–196. Available from <https://doi.org/10.1177/0361684316674721>.
- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *75*, 811–832. Available from <https://doi.org/10.1177/0146167205275304>.
- Plant, E. A., & Devine, P. G. (2001). Responses to other-imposed pro- Black pressure: Acceptance or backlash? *Journal of Experimental Social Psychology*, *37*, 486–501. Available from <https://doi.org/10.1006/jesp.2001.1478>.
- Plant, E. A., & Devine, P. G. (2003). The antecedents and implications of interracial anxiety. *Personality and Social Psychology Bulletin*, *29*, 790–801. Available from <https://doi.org/10.1177/0146167203029006011>.
- Plant, E. A., & Devine, P. G. (2009). The active control of prejudice: Unpacking the intentions guiding control efforts. *Journal of Personality and Social Psychology*, *96*, 640–652. Available from <https://doi.org/10.1037/a0012960>.
- Plaut, V. C., Thomas, K. M., & Goren, M. J. (2009). Is multiculturalism or colorblindness better for minorities? *Psychological Science*, *20*, 444–446. Available from <https://doi.org/10.1111/j.1467-9280.2009.02318.x>.
- Plaut, V. C., Thomas, K. M., Hurd, K., & Romano, C. A. (2018). Do color blindness and multiculturalism remedy or foster discrimination and racism? *Current Directions in Psychological Science*, *27*, 200–206. Available from <https://doi.org/10.1177/0963721418766068>.
- Prochaska, J. O., & Velicer, W. F. (1997). The transtheoretical model of health behavior change. *American Journal of Health Promotion*, *12*(1), 38–48. Available from <https://doi.org/10.4278/0890-1171-12.1.38>.
- Rattan, A., & Dweck, C. (2010). Who confronts prejudice? The role of implicit theories in the motivation to confront prejudice. *Psychological Science*, *21*, 952–959. Available from <https://doi.org/10.1177/0956797610374740>.
- Richeson, J. A., & Nussbaum, R. J. (2004). The impact of multiculturalism versus colorblindness on racial bias. *Journal of Experimental Social Psychology*, *40*, 417–423. Available from <https://doi.org/10.1016/j.jesp.2003.09.002>.
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt Rinehart Winston.
- Shelton, J. N., Richeson, J. A., Salvatore, J., & Hill, D. M. (2006). Silence is not golden: The intrapersonal consequences of not confronting prejudice. In S. Levin, & C. van Laar (Eds.), *Stigma and group inequality: Social psychological perspectives*. Mahwah, NJ: Erlbaum.
- Stephan, W. G., & Stephan, C. W. (1985). Intergroup anxiety. *Journal of Social Issues*, *41* (3), 157–175. Available from <https://doi.org/10.1111/j.1540-4560.1985.tb01134.x>.
- Stephan, W. G., & Stephan, C. W. (2001). *Improving intergroup relations*. Thousand Oaks, CA: Sage.
- Tropp, L. R., & Barlow, F. K. (2018). Making advantaged racial groups care about inequality: Intergroup contact as a route to psychological investment. *Current Directions in Psychological Science*, *27*, 194–199. Available from <https://doi.org/10.1177/0963721417743282>.

- Uhlmann, E. L., & Cohen, G. L. (2005). Constructed criteria: Redefining merit to justify discrimination. *Psychological Science, 16*(6), 474–480.
- Uhlmann, E. L., & Cohen, G. L. (2007). “I think it, therefore it’s true”: Effects of self-perceived objectivity on hiring discrimination. *Organizational Behavior and Human Decision Processes, 104*, 207–223. Available from <https://doi.org/10.1016/j.obhdp.2007.07.001>.
- Wakabayashi, D. (2017, September 08). *At Google, employee-led effort finds men are paid more than women*. Retrieved from <https://www.nytimes.com/2017/09/08/technology/google-salaries-gender-disparity.html>.