

Developing Scientifically Validated Bias and Diversity Trainings that Work:

Empowering Agents of Change to Reduce Bias, Create Inclusion, and Promote Equity

William T. L. Cox<sup>1,2</sup>

<sup>1</sup>Department of Psychology, University of Wisconsin – Madison

<sup>2</sup>Inequity Agents of Change Foundation

*Preprint, January 2022*

Author note

William T. L. Cox, Department of Psychology, University of Wisconsin – Madison, and Inequity Agents of Change Foundation.

Address correspondence to William Cox, Psychology Department, University of Wisconsin – Madison, 706 Vernon Ave, Madison, WI, 53714. Email: will.cox@me.com.

**Acknowledgements**

The preparation of this manuscript was supported by Grant 1R35GM128888 from NIGMS at the NIH. I'd like to thank Shaniya Auxier, Eric Roman Beining, Shannon Carnahan, and Curtis Ryals for their feedback and comments on this manuscript. I'd also like to thank Xizhou Xie for conducting the analyses reported in Figure 1, and creating Figure 1's Bayes factor graph. I also thank Eric Roman Beining for creating the bar graph in Figure 1. Last and far from least, I'd like to thank Trish Devine, my longtime friend and collaborator, without whom the work reported herein would not be possible.

## Developing Scientifically Validated Bias and Diversity Trainings that Work: Empowering Agents of Change to Reduce Bias, Create Inclusion, and Promote Equity

### Abstract

#### **Design/methodology/approach**

This viewpoint article draws primarily on review and analysis of the scientific literature related to bias, diversity, equity, and inclusion (DEI) trainings and cognitive-behavioral change, but also draws on firsthand experiences as a DEI practitioner. I provide a roadmap for adapting effective, scientifically validated approaches from other disciplines into the DEI context. Lastly, I review 15 years of randomized-controlled experiments testing the bias habit-breaking training, as one test case demonstrating the potential for DEI trainings that empower people to reduce bias, create inclusion, and promote equity.

#### **Purpose**

Research consistently shows that non-scientific bias, equity, and diversity trainings do not work, and often make bias and diversity problems worse. In this article, I argue that, despite widespread failures of the DEI industry, there is considerable reason to have hope that effective, meaningful DEI efforts can be developed.

#### **Findings**

DEI trainings fail due to widespread adoption of the information deficit model, which is well-known as a highly ineffective approach. Empowerment-based approaches, on the other hand are highly promising for making meaningful, lasting changes in the DEI realm. Evidence reviewed related to the bias habit-breaking training suggests that it empowers individuals as agents of change to reduce bias, create inclusion, and promote equity, both within themselves and the social contexts they inhabit.

#### **Originality/value**

In contrast to the considerable despair and pessimism around DEI efforts, the present analysis provides an empirically-validated path forward, to develop and test DEI approaches that empower individuals as agents of change.

## Developing Scientifically Validated Bias and Diversity Trainings that Work: Empowering Agents of Change to Reduce Bias, Create Inclusion, and Promote Equity

The last several years have seen a renewed, energetic attention toward social justice issues related to race, gender, and other historically oppressed groups. As more people recognize the need to address individual and institutional forms of bias within various aspects of their lives, there has been an increase in requests for workplace trainings and interventions related to bias, diversity, equity, and inclusion (DEI). Correspondingly, DEI trainings have proliferated under a variety of labels (e.g., “Diversity Training,” “DEI training,” “Sensitivity Training,” “Cultural Competency Training,” “Anti-Racism Training”). For simplicity, this article will use the term “DEI training” to encompass all of these types of training programs meant to create positive changes related to bias, diversity, equity, and inclusion.

Choosing from the broad array of DEI trainings can be difficult for managers or human-resources departments considering a training for their workplace. How does one know a training is effective? How can they be sure it will teach concepts that are relevant and applicable in their particular industry? Will a training help accomplish company goals? Finding answers to these questions is not a simple task, and navigating these issues is often frustrating. When it exists, the systematic research on current DEI trainings is clear, and unfortunate: at best they do not work, and at worst they exacerbate bias and diversity problems (for excellent reviews, see al-Gharbi, 2020; Devine & Ash, 2021; Paluck et al., 2021; Pendry et al., 2007). Mainstream awareness of the issues with corporate and academic DEI trainings has increased, with major news outlets, including *Scientific American*, *Forbes*, *Business Insider*, *TIME*, and many others releasing articles

denouncing that DEI trainings do not work (Green & Hagiwara, 2020; King, 2020; Lipman, 2018; Tang & Huang, 2021). This state of affairs is undoubtedly very frustrating for individuals or institutions interested in making meaningful changes related to bias and diversity. Is there any hope for those who want to reduce bias, create inclusion, and promote equity in their organizations?

In the present article, I argue that there is indeed hope — the abundant failures of the DEI industry arise from a root cause that is well-understood and has been identified and overcome in many other content areas. As such, we can learn from efforts in other content areas how to develop approaches that create positive changes. Specifically, most DEI trainings adopt *information deficit model* approaches, which assume recipients lack key information, then try to correct that deficiency. Often used as an intuitive first attempt at creating change, the information deficit model's shortcomings are well-documented, as are alternative approaches that more effectively create change. In contrast to deficit models that treat people as passive recipients of information, *empowerment-based* approaches respect people's autonomy, and equip them to be effective, self-motivated agents of change within themselves and within their institutions. I discuss insights into how these kinds of approaches can address key challenges faced by DEI trainings, and, finally, review 15 years of experimental evidence demonstrating the efficacy of one particular empowerment-based DEI training, the bias habit-breaking training. In contrast to standard, non-scientific approaches that have proliferated in the public domain, the bias habit-training has shown considerable promise at creating lasting, meaningful changes, both within individuals and institutions.

The present article draws primarily on scientific evidence and analysis, drawn from my background as an empirical researcher in the realm of stereotyping, bias, and evidence-based cognitive-behavioral change. I also sometimes draw on my 20 years of experience working in the DEI realm: I have worked with many organizations to develop and implement DEI initiatives and, specifically, to administer the bias habit-breaking training. Working with diverse workplaces and industries has given me a number of insights into specific practical challenges faced by DEI trainings in the world, and how empowerment-based approaches are uniquely suited to addressing these challenges.

### **Standard DEI Trainings are Just the Latest Failures of the Information Deficit Model**

The state of DEI trainings currently being implemented in the world, including their common content, approaches, and assessments of their efficacy (if any) has been extensively reviewed elsewhere. Rather than duplicating the excellent work of other authors here, I will briefly state that the overwhelming consensus is that 1) DEI trainings out in the world are largely non-scientific and not experimentally tested, 2) the limited but consistent non-experimental evidence suggests that at worst they cause more problems for organizational climate, and at best they are ineffective at creating lasting, meaningful change, and 3) experimental tests of their common components lead to more negative outcomes than positive outcomes related to bias, diversity, inclusion, and equity (for excellent reviews, see al-Gharbi, 2020; Devine & Ash, 2021; Paluck et al., 2021; Pendry et al., 2007; see also Apfelbaum et al., 2012; Bezrukova et al. 2016; Brady et al., 2015; Brannon et al., 2018; Brewer et al., 1999; Byrnes & Kiger, 1990; Carter et al., 2020; Cooley et al., 2019; Cox & Devine, 2019; 2022; Dobbin & Kalev, 2013; 2016; Dobbin et al., 2007; 2015; Dover et al., 2014; Duguid & Thomas-Hunt, 2015; FitzGerald et al., 2019; Frisby, &

O'Donohue, 2018; Folz, 2016; Kalev et al., 2006; Kaiser et al., 2013; Kulick et al., 2000; Legault et al., 2011; Lilienfeld, 2017; Lipman, 2018; Magley et al., 2016; Moss-Racusin, et al., 2014; Naff & Kellough, 2007; Newkirk, 2019; Nordell, 2017; 2021; Paluck, 2012; Paluck & Green, 2009; Rynes & Benson, 1995; Siden et al., 2021; Stephan & Stephan, 2001; Stewart et al., 2003; Wilton et al. 2019).

The lack of evidence in favor of DEI trainings, coupled with the abundance of evidence against them, has led many to directly question the ethics of implementing untested, non-scientific DEI trainings (e.g., al-Gharbi, 2020; Cox & Devine, 2019; Dobbin & Kalev, 2013, 2016; Nordell, 2017; Paluck, 2012). Nearly every major scientific organization (e.g., NIH, NSF, AAAS) has called for experimentally-tested, evidence-based approaches to addressing bias and promoting diversity (e.g., Moss-Racusin, et al., 2014). Paluck powerfully argued that we should consider experimental testing of DEI trainings to be an “ethical imperative” (Paluck, 2012). The DEI trainings are estimated to be an \$8 billion industry (Carter et al., 2018; Lipman, 2018), and some have gone so far as to declare it a *failed* industry (Newark, 2019).

I argue that the pervasive failures of the DEI industry have close, fundamental ties to failures at creating change in other content areas. Specifically, DEI trainings overwhelmingly adopt an *information deficit model*, which has long been identified as a ubiquitous but highly ineffective approach to creating cognitive-behavioral change (Boykoff, 2011; Dickson, 2000; Irwin & Wynne, 1996; McDivitt, 2016; Miller, 2001; Schultz, 2002; Suldovsky, 2017). When seeking to change behavior in an audience (e.g., the lay public, students, employees), the first impulse of many communicators is to educate that audience — to give them information they are perceived

to lack. The assumption inherent to this approach is that people fail to engage in desired behaviors (or persist in undesired behaviors) due to a knowledge deficit. Therefore, correcting that deficiency is the apparent and obvious path to change (Boykoff, 2011; Dickson, 2000; 2005; McDivitt, 2016; Miller, 2001; Schultz, 2002; Suldovsky, 2017). A dermatologist, for example, might assume that people fail to wear sunscreen because they do not understand their risk and the dangers of skin cancer. The dermatologist then believes that teaching people about how sunscreen can help prevent skin cancer will correct the knowledge deficit and therefore be effective at getting people to wear sunscreen. Often, providing this information is the end of the dermatologist's involvement, because their proximal goal — to correct the information deficit — has been met. When deficit model approaches like this are experimentally tested, however, the net effect is most often either no change or *decreases* in the desired cognitive or behavioral outcome (in the case of sunscreen use, see Jensen et al., 2020; Kantor, 2020).

Importantly, those employing deficit model approaches do not necessarily do so *consciously* (Suldovsky, 2017). Rather, the deficit model is a default approach because it is intuitively appealing and seems as though it should be effective — after all, if lack of information is the problem, then providing more information is the obvious solution. The intuitive appeal of the deficit model has made it ubiquitous, even though its documented failures are legion (e.g., in the realms of smoking behaviors, McCarthy, 1985; Tønnesen, 2002; climate change McDivitt, 2016; Schultz, 2002; Suldovsky, 2017; and vaccines, Dubé et al., 2013; 2015; Hornsey et al., 2018; Jarrett et al., 2015; Rossen et al., 2016; Sadaf et al., 2013; Smith, 2017; Trevors & Duffy, 2020). Despite widespread criticism, abundant evidence of its failure, and even being officially discredited by scientific bodies, the deficit model repeatedly resurfaces in a variety of new forms

(Boykoff, 2011; Dickson, 2000; Kearnes et al., 2006; McDivitt, 2016; Miller, 2001; Suldovsky, 2017).

The clear trend of deficit model approaches is that they either do nothing or make problems worse — a trend clearly echoed by DEI trainings. Efforts in the DEI context, therefore, can benefit from lessons learned in other content areas, to move away from the default of the information deficit model, and toward more effective, scientifically validated models of change. The information deficit model adopts a paternalistic, top-down approach that treats people as passive processors of information, and it tacitly assumes that behavior is determined by a single predictive factor: the presence or absence of relevant knowledge (Marteau et al., 1998; McDivitt, 2016; Miller, 2001; Simis et al., 2016). The stated or inferred foundation of these approaches is that there is something “wrong” with the recipients, which is predestined to make people feel defensive and to decrease motivation to change (i.e., it creates reactance; Brehm & Brehm, 2013). Further, people’s behavior is determined by a complex interplay of factors, including not only their knowledge, but their values, motivations, goals, sensitivity to social pressures, anxieties, and many other social and psychological forces (Boykoff, 2011; Cox, 2015; Dickson, 2000; Kearnes et al., 2006; McDivitt, 2016; Suldovsky, 2017; Wynne, 1992). Creating meaningful, lasting changes in a given content area requires not only expertise in that area, but expertise in the science of cognitive-behavioral change and intervention science.

To my knowledge, no prior scholar has explicitly named the information deficit model as the root cause of the widespread failure of the DEI industry, but it becomes readily apparent when reading reviews that comprehensively document common DEI training methods and activi-



ties (e.g., Carter et al., 2020; Devine & Ash, 2021; Paluck & Green, 2009; Paluck et al., 2021; Pendry et al., 2007). As reviewed extensively by Pendry and colleagues (2007), very common DEI training content involves “raising awareness” of historical intergroup disparities, the existence of biases in modern society or the workplace, implicit bias, White privilege (McIntosh, 1988), microaggressions (Lilienfeld, 2017), how it feels to be excluded/discriminated against (Byrnes & Kiger, 1990; Stewart et al., 2003), or the benefits of diverse workforces. These reviews converge with my firsthand experience working in the DEI realm — DEI efforts I have seen, heard about, or been asked to comment on almost invariably fall neatly into the information deficit model. As in other contexts, the information deficit model is predictably ubiquitous in DEI. Common DEI trainings 1) assume bias and equity issues arise from deficits in employees’ knowledge of biases, racism, sexism, and other DEI issues, and 2) believe that providing information or experiences to ‘correct’ these deficits will lead to meaningful cognitive-behavioral change. Organizational DEI efforts are just the latest example of the insufficiency of the information deficit model to effect meaningful, lasting change.

### **A Shift in Approach: Empowering Agents of Change**

A core flaw of deficit model approaches is that they fail to enlist recipients as active, autonomous agents in the change process. These approaches fail to recognize people’s inherent better nature and make the mistake of trying to impose change on recipients. Dickson (2000) previously argued that a clear remedy to deficit model approaches is to adopt empowerment-based approaches. The foundation of empowerment-based approaches is to center individuals as the primary drivers of the change process (see Cox & Devine, 2019). Rather than change being something that is done *to* people, empowerment-based approaches focus on working *with* people,

and empowering them as agents of change (see also Hennes et al., 2018). Contrasted with the deficit model's notion that information is the sole determinant of behavior, empowerment-based approaches engage dynamically with people's preexisting values, motivations, social connections, and other psychological processes. For an empowerment-based DEI approach, the starting point is to respect people's autonomy and believe that they have good intentions with regard to bias and diversity issues (Cox & Devine, 2019). Although empowerment-based approaches may teach people new information, this education is framed as *enhancing* people's existing knowledge and experiences, rather than correcting a deficit.

This shift from deficit to empowerment approaches has parallels in many other domains that have recognized the benefit of shifting from paternalism to autonomy (Cook, 2001). In social and behavioral sciences, a strong parallel is the paradigm shift from treating and calling people "subjects" of the research, to considering people to be willing and informed "participants" (e.g., Boynton, 1998; Sales & Folkman, 2000). Recent medical scholars have made the case for transitioning away from the traditional medical model, in which medical professionals give directives that patients are expected to obey, to models like patient-centered care, in which medical professionals strive to give patients the tools to be informed decision-makers and then work *with* the patients to develop treatment plans (e.g., Kumar & Chattu, 2018). Similar discussions in management science have arisen in recent years, emphasizing transformational leadership over transactional and directive leadership styles (Arnold & Loughlin, 2013; Odumeru & Ogbonna, 2013), or employee-centered/human-centered management as opposed to mechanistic management (Baker et al., 1996; Hoogervorst et al., 2005). In different contexts, these and other recent conversations have all begun to emphasize the benefits of moving away from paternalistic ap-

proaches that tell people what to do and toward approaches that respect people's autonomy and give them tools to direct their own actions toward shared goals (see also Howe et al., 2022).

The success of an empowerment-based approach requires that people's autonomous, intrinsic values and motivations are compatible with the overarching goals of the training. Fortunately, basic research has long demonstrated considerable reasons for optimism about approaches that trust people's autonomy around DEI topics. Most people hold strong personal values that oppose bias, prejudice, and inequity — they want to treat others fairly, unencumbered by biases or prejudgments (e.g., Cox, 2015; Devine, 1989, Devine et al., 1991; Monteith, 1993; Monteith et al., 2001; 2002; Plant & Devine, 1998; 2009). Since the 1980s, research has shown repeatedly that the average person believes in fairness, and will become motivated to reduce bias and inequity if they are made aware that they may, however unwittingly, be vulnerable to biases (Amodio et al., 2007; Cox, 2015; Devine, 1989, Devine et al., 1991; Monteith, 1993; Monteith et al., 2001; 2002; Plant & Devine, 1998; 2009).

Motivation to address bias and inequity, however, is not sufficient — very often, people direct their motivation in unhelpful directions, adopting intuitive tactics that seem like they *should* help reduce bias but that actually backfire, or are simply inadequate (e.g., Apfelbaum et al., 2008; Norton et al., 2006; Uhlmann & Cohen, 2007). Trying to suppress stereotypes, for example, leads a rebound effect where people display *more* bias, rather than less (Macrae et al., 1994). The focus of an empowerment-based approach, therefore, is to engage people's pre-existing personal values and motivations related to DEI, guide them away from common ineffective tactics, and help them to autonomously direct their efforts in more effective directions. To identi-

fy effective *methods* to guide the change process in this way, one can look to other areas of work that have developed and successfully implemented models of self-sustaining cognitive-behavioral change.

### **Scientific Methods to Effect Lasting Cognitive-Behavioral Change**

Any effort to effect meaningful changes in people, whether related to bias and diversity or any other context, should make use of the vast literature on the science of cognitive-behavioral change. In fact, many scientists have long argued that, rather than treating intergroup biases as a “special” or fundamentally distinct type of psychological phenomena, they should instead be understood as arising from ordinary psychological processes (Allport, 1954; Bodenhausen & Macrae, 1998; Cox & Devine, 2015; Cox et al., 2012; Devine, 1989; Fiske, 1998). The social *impact* of biases may be special, but their underlying cognitive origins are not. Efforts to create lasting, meaningful cognitive-behavioral change related to DEI, therefore, can build on effective approaches from other content areas.

### **The Habit Model**

Devine (1989) has long argued that intergroup biases can be likened to habits of mind. From the media and their social environments, kids learn stereotypes about major social groups and display biases based on those stereotypes at ages as young as 4, 5, and 6 years old (Bigler & Liben, 2007; Levy & Killen, 2008; Pauker et al., 2010). These stereotypes and biases are reinforced over the lifetime, making them automatic, default responses that are often at odds with conscious values and intentions — just like bad habits (Devine, 1989). As adults, we have the cognitive reasoning capacity to regulate these habitual stereotypes and biases, but we continue to

be bombarded by stereotypes in the media, and many processes in our cognitive systems (e.g., confirmation bias; illusory correlation) help perpetuate these stereotypes in our minds (Chapman, 1967; Cox et al., 2012; Darley & Gross, 2004; Gibson et al., 2013; Kalish et al., 2011; Nickerson, 1998; Pohl, 2004). Biases, like other habits, are automatic, persistent, and often operate at odds to conscious intentions. If we consider biases as habits of mind, one potentially fruitful approach to changing biases is to draw on the habit change literature (e.g., see Wood, 2017; Wood & Neal, 2016) and to approach bias reduction as a process of breaking a habit (Cox & Devine, 2019; Devine, 1989; Devine et al., 2012).

Considering the notion of biases as habitual highlights a major way empowerment-based approaches can succeed where the information deficit model fails. A core flaw of the deficit model is the tacit assumption that persistence of undesired behaviors (or lack of desired behaviors) occurs due to a passive *lack* of information. Habits, however, are maintained by interacting, active forces for inertia. Combatting this inertia requires active, sustained effort over time. Empowerment-based approaches are well-suited to imparting this message, if they give people *actionable* tools to employ in the service of combatting biases and help make the change process *self-sustaining* over time. Approaching bias and inequity reduction using a “habit” model is especially useful for addressing this challenge, because people intuitively understand that habits have considerable inertia, and inherent to the notion of breaking a habit is that it requires sustained effort over time (Cox & Devine, 2019).

**Cognitive Behavioral Therapy**

In many ways, stereotypes and biases are directly parallel to other types of “habits of mind”, especially negative self-schemas that have been of interest to clinical researchers (Cox et al., 2012). Just as stereotypes are automatically activated cognitions that negatively affect judgments, feelings, and behaviors toward others, negative self-schemas in depression are automatically activated cognitions that negatively affect judgments, feelings, and behaviors toward the self. Both intergroup stereotypes and negative self-schemas are well-learned, well-rehearsed cognitive structures that are automatically activated, are difficult to change, can bias attention and information processing, and have an array of cognitive, affective, and behavioral consequences that are often at odds with conscious intentions (see Cox et al., 2012 for a more comprehensive review; Bargh & Tota, 1988; Beck, 1967; Beck & Alford, 2009; Devine, 1989; Dunn & Spellman, 2003; Eaves & Rush, 1984; Fiske, 1998; Hamilton & Trolie, 1986; Hilton & von Hippel, 1996; Wenzlaff, Wegner, & Roper, 1988). Building on this insight, Cox and colleagues (2012) argued that methods from experimental clinical research, especially cognitive-behavioral therapy (CBT), could be especially useful in approaches to address intergroup biases (see also Beck, 1999).

CBT is one of the oldest and most widely-applied behavioral change frameworks, and it has been extensively validated in decades of experimental studies (for reviews, see Cuijpers et al., 2016; Hofmann et al., 2012; 2013). CBT enlists the participant (or patient) as an active agent of their own change, helping them to identify maladaptive cognitions and behaviors and giving them concrete cognitive and behavioral tools to help them change the maladaptive cognitions/behaviors. CBT for depression, for example, is as efficacious as medication for depression, and

also reduces likelihood of relapse after treatment ends, because CBT equips people to continue their therapeutic work themselves (Beck, 2005; Beck, 2021; Hollon, 2003; Hollon & Dimidjian, 2009; Hollon & Shelton, 2001). In the clinical realm, CBT is, by far, the most effective method at creating long-term change in individuals' cognitive-behavioral processes and their affective consequences (Beck, 2005; Beck, 2021; Hollon, 2003; Hollon & Beck, 2013; Hollon & Dimidjian, 2009; Hollon & Shelton, 2001). CBT has been extensively studied and carefully refined since its origination in 1970s, and therefore has precise, established methods and parameters for effectively guiding long-lasting cognitive-behavioral change (e.g., Beck, 2005; Beck, 2021; Cuijpers et al., 2016; Hofmann et al., 2012; 2013). Some work even shows long-lasting beneficial effects from just a single session (e.g., Schleider & Weisz, 2018), or from CBT delivered via a computer application (e.g., Carroll et al., 2008; Cavanagh et al., 2006; Ebert et al., 2015; Himle et al., 2006; Luo et al., 2020). Crucial to CBT is that clients are taught how to recognize, manage, and address automatic cognitions and behaviors autonomously — CBT gives them tools to make their change process self-sustaining over time (Beck, 2021). Cox and colleagues (2012) developed an integrated theoretical model adapting CBT methods as a self-sustaining approach to reduce bias. Drawing on well-established, effective methods for cognitive-behavioral change provides a clear blueprint for guiding change in a DEI context.

The versatility of CBT is especially useful given that there are infinite forms of diversity, and therefore bias, to consider. People are vulnerable to showing biases related to race, gender, sexual orientation, political orientation, religion, disability, many other social identities, and the intersections thereof (Cox et al., 2012). It is infeasible for organizations to implement separate interventions or training programs for every stereotyped group that may be affected by bias (e.g.

Black people, Latin people, Asian people, Muslim people, women, LGBT people, people with disabilities). In addition to the diversity of potential target groups, bias and inequity take diverse forms, across and within individual people, time periods, circumstances, and institutions (Cox et al., 2012). Considering the various target groups and the various forms bias and inequities take, the variability of bias is truly limitless. No training approach could hope to comprehensively tally them all.

Considering the infinite variability of biases highlights another crucial way that information deficit model approaches fail. Because the deficit model relies on directive information transfer, audience learning is largely limited to, at most, a “laundry list” of facts, which, in the case of biases/inequities, will inevitably be incomplete. Empowerment-based approaches, on the other hand, seek to teach recipients skills that enable them to autonomously recognize and address novel issues that may arise. Insights from CBT are especially relevant to addressing this challenge: A therapist cannot anticipate all the specific forms their client’s myriad stressors and life circumstances will take in the future, so the therapist must teach the client *generalizable* and *customizable* skills (Beck, 2021). A DEI training, correspondingly, must likewise teach training recipients generalizable and customizable skills to identify and address the infinite variability of biases and inequities they may encounter. By centering the individual, not the trainer, as the primary agent of change, empowerment-based approaches provide a model to address the infinite varieties of bias and inequity.



### **The Bias Habit-Breaking Training**

One empowerment-based approach that encapsulates the principles, methods, and challenges reviewed above is the bias habit-breaking training (also sometimes known as the “prejudice habit-breaking intervention” or the “break the bias habit” workshop; Carnes et al., 2015; Devine et al., 2012; Cox et al., 2022; for a review see Cox & Devine, 2019). Since 2007, this training was developed as an approach to empower people to be autonomous agents of change. In this context, agents of change are defined as individuals who are self-motivated and actively engaged in reducing bias, creating inclusion, and promoting equity, both within themselves and within the institutions and social systems they interact with. The training is designed to help them sustain these efforts across their lifetimes, equipping them to seek out and address new forms of bias whenever and wherever they arise.

The intervention was developed, experimentally tested, and iteratively refined and updated over the past 15 years. Common across all versions of this approach is that participants are taught Devine’s (1989) prejudice habit model, which discusses biases and stereotypes as “habits of mind” and describes the process of overcoming these biases as breaking a habit (Devine, 1989). Breaking a habit is an intuitive, familiar idea for people, making it easy to apply in the context of addressing biases. Core to the habit-breaking approach is that breaking a habit requires ongoing effort over time, setting people up to understand that they must make their efforts self-sustaining. Of key interest is relaying the intervention content in a way that will maximize recipients’ retention of the information, motivation to work on DEI issues, and likelihood of sustaining these efforts over time. Building on Cox and colleagues’ (2012) framework, the bias

habit-breaking training adapts core principles from CBT to help meet these goals (Beck, 2021; see Table 1).

**Table 1.** Beck's (2021) 10 Core Principles of CBT, and their parallels in the bias habit-breaking training

<b>CBT...</b>	<b>The Bias Habit-Breaking Training...</b>
Is based on an ever-evolving formulation of patient and their problems in cognitive terms	Teaches participants that bias unfolds in myriad ways, and gives them concrete cognitive terminology to understand various forms of bias
Requires a solid relationship between patient and therapist	Is delivered by an expert presenter, with both deep content expertise and skill developing rapport with participants
Emphasizes collaboration and active participation	Empowers participants to operate as autonomous the agents of change
Is goal-oriented and problem-focused	Orients participants to specific, actionable steps they can take to make change
Initially emphasizes the present	Starts participants focusing on what they can influence most in the present, providing a foundation to build on
Is educative, teaching the client to be their own therapist and emphasizing relapse prevention	Directly teaches participants how to continue applying the cognitive-behavioral change process, sustaining it into the future
Aims to be time-limited	Is designed to give participants the fundamentals needed in a single session
Has carefully structured sessions	Has a carefully structured format designed to maximize motivation and information retention
Teaches patients to identify, evaluate, and respond to their dysfunctional thoughts and beliefs	Attunes participants to the key ways cognitive biases lead to bias and inequity, and teaches them to disrupt them
Uses a variety of techniques to change thinking, mood, and behavior	Teaches a variety of tools and skills to reduce bias, create inclusion, and promote equity

The intervention content teaches people how and *why* biases occur, encouraging them to autonomously seek out and identify the variable forms of bias that may occur in their own lives. It also teaches them actionable, generalizable, customizable tools that they can use, if they so choose, to reduce bias, create inclusion, and promote equity, related to any target group. There are a number of tools, including tools that involve retraining cognitive/emotional reactions to reduce bias, procedures that can be put into place to prevent bias, ways to create more inclusive environments, how to effectively speak up about bias, and other topics. Integral to the training is that bias is discussed in a very evidence-based, non-accusatory way — it frames the scientific evidence as demonstrating people’s *vulnerability* to biases, rather than presenting bias as inevitable or ubiquitous. Each component of the training is carefully crafted to directly address challenges created by biases, and to help recipients be motivated, effective, autonomous agents of change.

### **Evidence of Effectiveness/Efficacy**

Several experimental studies have demonstrated lasting and impactful effects of the bias habit-breaking training. As noted earlier, the training has been iteratively refined and updated over the last 15 years. The very early versions of the intervention focused more narrowly on one type of intergroup bias (e.g., race bias, Devine et al., 2012; Forscher et al., 2017; gender bias, Carnes et al., 2015). Following the initial success of these early versions, the training was updated to cover *general* principles of how biases can play out related to any sort of target group and in any situation or context (e.g., Cox et al., 2022). The core components of the training, as described above, are largely constant across these versions. I organize the review of these findings under three broad themes that correspond to key areas where DEI programs seek to make

progress: Reducing Bias, Creating Inclusion, and Promoting Equity. Following this review of published and ongoing quantitative, empirical research testing the bias habit-breaking training, I also share a few (admittedly anecdotal) examples of impacts the training has had at various organizations that showcase the breadth of potential outcomes that can be influenced when individuals are situated and empowered as the primary agents of change.

### ***Reducing Bias***

In the very first test (Devine et al., 2012) of the first version of the habit-breaking training, 91 undergraduate students completed an array of baseline measures, then were randomly assigned to serve as controls or to receive the training. They completed follow-up assessments at several timepoints, up to 8 weeks post-manipulation. The first outcomes of interest were participants' self-reported levels of awareness of their vulnerability to express bias unintentionally and their concern that racial bias was a serious problem (see also Carter et al., 2020). Whereas control participants' scores on these measures remained unchanged over time, training participants' scores significantly increased, indicating the training was effective at making people more aware of their potential to express race bias unintentionally, and more concerned that it was a serious problem. These effects were later replicated in two other, higher-powered randomized controlled trials (Cox et al., 2022; Forscher et al., 2017).

In the initial randomized-controlled trial (Devine et al., 2012), training participants' levels of implicit bias, as measured by the Implicit Association Test (IAT), significantly decreased, whereas control participants' IAT scores did not. This effect endured to the end of the study, 8 weeks post-manipulation. This effect was unprecedented in the literature at the time, and it re-

mains so today. In a recent meta-analysis of 492 experiments trying to reduce implicit bias (Forscher, Lai, et al., 2019), no study demonstrated decreases in implicit bias that lasted more than 24 hours, with most lasting only a few minutes (see also Lai et al., 2016; Siden et al., 2021). Work on the bias habit-breaking training is the sole exception to these patterns. In fact, the bias habit-breaking training's approach is so substantially different from other interventions that Forscher, Lai, and colleagues excluded it from their meta-analysis. Devine and colleagues' (2012) pattern of reduced implicit bias has recently been replicated in a new randomized-controlled experiment with a much larger sample size ( $N = 957$ ). In this replication (Cox et al., 2022), training participants again significantly decreased in implicit bias, whereas control participants' IAT scores remained unchanged. This effect was observed up to the latest timepoint during which IAT scores were collected, at 6 weeks post-manipulation.

The importance of a reduction in measured implicit bias lies in the assumption that this reduction will correspond to reductions in other outcomes; the IAT is very often used as a proxy for discriminatory judgments and behaviors (Cox & Devine, 2022). In the Forscher, Lai, et al. (2019) meta-analysis, when studies assessed implicit bias interventions alongside behavioral outcomes, observed decreases in IAT scores did not mediate corresponding reductions in biased behavior. This lack of mediation indicates that those reductions in measured implicit bias are unlikely to be meaningful for other outcomes (Forscher, Lai, et al., 2019). Again, work with the bias habit-breaking training appears to be the sole exception. A subsample ( $N = 320$ ) of Cox and colleagues' (2022) participants were recruited for an ostensibly unrelated study 2-3 years after being randomly assigned to receive the bias habit-breaking training or to serve as controls. Participants completed Monteith and colleagues' (2002) stereotype regulation task. This measure is

commonly used as an indicator of the amount of effort people will put into avoiding stereotypic assumptions (see Burns et al., 2017; Czopp et al., 2006). Training participants stereotyped less than controls on this task, and this effect was significantly mediated by the prior observed decrease in implicit bias (Cox et al., 2022). Unlike other documented reductions in implicit bias (Forscher, Lai, et al., 2019), the reductions resulting from the bias habit-breaking training appear to be both long-lasting and related to other, meaningful outcomes.

### ***Creating Inclusion***

In addition to reducing bias, a common goal of DEI initiatives is to encourage behaviors that create more inclusive environments. One important way that people engage in inclusivity is to confront biased rhetoric in others, and to speak up in favor of members of stigmatized groups. Testing the training's effects on speaking up, Forscher and colleagues (2017) randomly assigned 302 undergraduate participants to serve as controls or to complete the habit-breaking training. In an ostensibly unrelated study 1-2 years post-manipulation, participants received an email that they believed was sent by their school newspaper. The email asked them to be part of a program in the newspaper that allowed a student to write an editorial about a topic of their choosing, which was then reviewed by fellow students who could, if they so chose, write a reply that would be published alongside the editorial. The (fabricated) editorial participants read was titled "Racial stereotypes are useful tools", and it argued that stereotypes are useful, harmless, and that stereotyping has only become "untrendy" because society is too politically correct. Participants ( $N = 79$ ) rated their agreement/disagreement with the author's perspective, and they were given the option to write a response to the editorial, which they believed could be selected to be published under their name in the school paper. Training and control participants disagreed with the au-

thors' perspective to the same extent, but training participants were 64% more likely to translate that disagreement into behavior, taking an ostensibly public stand against the rhetoric by writing a response to the editorial that contradicted its authors' perspective (Forscher et al., 2017). These patterns provide further evidence for the value of an empowerment-based approach: the parity in training and control participants's disagreement with the bias rhetoric demonstrated again that most people's values oppose bias. The fact that training participants were more likely to speak up demonstrated that this training approach empowered them to translate those preexisting values into meaningful action.

Cox and colleagues (2022) likewise sought to examine speaking up about bias and inclusion related topics. A subsample ( $N = 304$ ) of their replication study described above completed an ostensibly unrelated study 1-2 years post-manipulation. This study involved a mock online classroom discussion, where students discussed popular press articles related to "hot topics" in contemporary discourse. Two articles touched on bias/inclusion-related issues, one discussing Muslim people being targeted for extra security screenings at airports, and one discussing the need for gender nonbinary bathrooms. These two DEI topics were not mentioned in the content of the training that the training participants had completed prior, thus enabling Cox and colleagues to examine whether training participants would generalize what they had learned beyond biases explicitly mentioned in the training. Quantitative text analyses assessed how much participants brought up bias and inclusion topics in their discussion of the Muslim and gender nonbinary articles. Compared to controls, training participants spoke up about bias/inclusion 181% more than control participants overall. For the Muslim-related article specifically, training participants discussed bias/inclusion topics 20.7% more than controls, and for the gender-nonbinary-

related article, training participants discussed bias/inclusion topics 12.5% more than controls. Each of these effects was statistically significant. This pattern is consistent with the prediction that this empowerment-based approach would give participants tools to generalize and customize what they learned in the training to additional, novel forms of bias not addressed directly within the training.

Another component of creating inclusion is to help others to be more inclusive. In response to open-ended questions about their experience with the training, many participants report sharing what they learned with their peers, to help more people create inclusive environments (Devine et al., 2012; Forscher et al., 2017). In both lab studies and real-world implementation of the bias habit-breaking training, Forscher (2017) demonstrated that people explicitly taught others how to reduce bias and create inclusion. In one study, the effects of the training were even stronger on people who did not attend the training, but who worked closely with someone who did attend the training (Forscher, 2017). This effect suggests that people not only share what they learned, but that as they do, they may customize the content even further, to have stronger effects in their specific context.

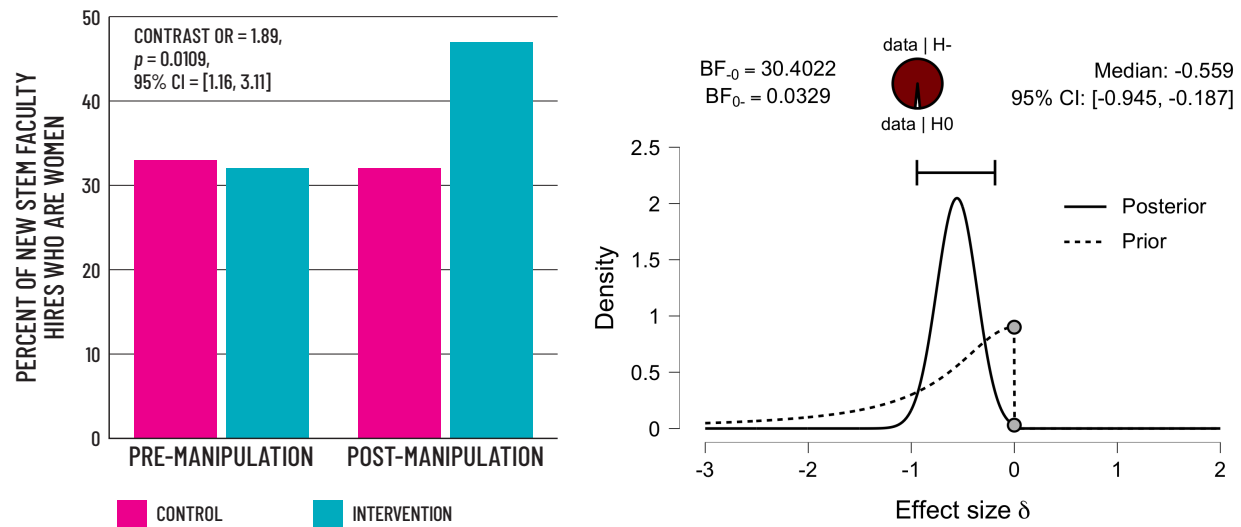
Some final indicators of creating inclusion relate to organizational climate. In another large-scale study, 92 academic departments in science, technology, engineering, and math (STEM) fields were randomly assigned to receive the bias habit-breaking training or to serve as controls (Carnes et al., 2015; Devine et al., 2017). This version of the training focused specifically on gender bias in STEM. Training faculty increased compared to controls on a number of self-report measures related to departmental climate, such as self-efficacy to address gender bias in



their department, having taken action to promote gender equity, and awareness of bias. Carnes and colleagues (2015) linked their study data with an office on campus that annually collects faculty climate data, to assess whether climate changed in training versus control departments. Whereas control department climate metrics remained the same over time, training department climate significantly improved over time, with both women and men in training departments reporting better departmental “fit” and that their work was more respected by their colleagues (Carnes et al., 2015).

### ***Promoting Equity***

Inequity can take many forms, in different contexts. In STEM contexts, a crucial issue is inequity in the hiring of women as tenure-track faculty. A recent, long-term follow-up to the Carnes et al. (2015) study examined whether the training led to changes in hiring of women as new faculty (Devine et al., 2017; Forscher, 2017). Working with human resources data, observed hiring patterns were consistent with expectations that women are underrepresented among new STEM hires: during the two years prior to the Carnes et al. (2015) study, new hires in both intervention and control departments were only 32-33% women. In the two years after random assignment, the control departments’ new hires were still only 32% women, while intervention departments’ new hires were 47% women, Contrast OR = 1.89,  $p = 0.0109$ . See Figure 1. In other words, the bias habit-breaking training caused a 43% increase in hiring of members of underrepresented groups.



**Figure 1. Hiring Data - Descriptives (Left) and Bayes Factor Analysis (Right)**

We conducted a Generalized Linear Mixed Effects Model with a logit link in the binomial family with the bobyqa optimizer in the lme4 package in R (Bates et al., 2015). The analysis included random intercepts for each departmental cluster and each matched pair of clusters (see further details about clusters in Devine et al., 2017). The number of women hires was weighted by the total number of hires within the cluster, so the outcomes could be interpreted as proportions. We tested the contrast comparing hires in intervention departments post-manipulation to hires in control departments pre- and post-manipulation and intervention departments pre-manipulation.

The proportion of women hired by intervention departments after the training was significantly higher than intervention departments before the training, or the control departments before or after the manipulation,  $OR = 1.89$ ,  $p = 0.0109$ , 95% CI = [1.16, 3.11].

Bayesian Independent Samples T-test (Right) of proportion of women hired by intervention departments after the training vs intervention departments before the training, or the control departments yields  $BF_{-0} = 30.40$ , indicating that the data are 30.40 times more likely to occur under the alternative hypothesis than the null hypothesis. This additional Bayesian test indicates there is *very strong* evidence that the bias training led to increased hiring of women.

These general hiring data were previously explored in Devine et al., (2017), but these figures and analyses are new for the present publication.

Other, ongoing work is examining additional outcomes related to important equity indicators in different contexts. An ongoing study with middle and high school teachers and students in a large Californian school district is examining whether teachers completing the habit-breaking training has beneficial effects on achievement gaps among their students. Preliminary data with 176 teachers and 7084 students indicate that when students have few teachers who received the habit-breaking training, there is a typical achievement gap, with Latin students having lower grades than White students. When students have a high proportion of teachers who received the training, however, that achievement gap disappears, driven by Latin students receiving higher grades (Saad et al., *in progress*). Lastly, in an ongoing study in Toronto, preliminary data indicate that doctors who were randomly assigned to receive the bias habit-breaking training received higher satisfaction marks from their Indigenous patients compared to doctors who were randomly assigned to serve as controls.

### ***Firsthand Examples***

As a scientist, I rely on the systematic, experimental data referenced above for my confidence that the bias habit-breaking training has beneficial, long-lasting effects. As a practitioner, I also have the privilege of witnessing firsthand how attendees engage with the training and apply it. My colleagues and I have delivered the training live for many companies, organizations, academic institutions, and governmental departments around the country and world. Over the last several years, I've given the training to dozens of audiences, in many different fields and contexts. Anecdotally, I can say that nearly every time the training is delivered, someone in the audience has what my colleagues and I call an "Aha! Moment", where they discover some way that they have been expressing bias without realizing it. Very often, these are highly specific forms of

bias that are particular to the individual or a particular workplace — examples unlikely to be captured in typical empirical, quantitative studies.

At one company that received the training, learning about how biases persist within organizational structures led an employee to investigate the method used by their company's mail-sorting machines, which stamped mail with an "L", or an "H" based on ZIP code. The current mailroom employees had no knowledge of why the mail sorting machine did this. It had been programmed to follow the manual sorting procedures of the previous mailroom staff, who had done what they were taught from the generation before, and so on. The account managers who received the sorted mail knew that "L" and "H" indicated low and high priority, respectively, but they did not know how those determinations were made (one account manager thought the priority status was determined by artificial intelligence software identifying known high-value clients). The manual sorting procedure was revealed to have been a holdover from the 1960s, when the company was explicitly discouraging its employees from working with Black clients — the mail marked as low priority came from the zip codes of historically Black neighborhoods. No one currently at the company knew the explicitly racist origins of these procedures, and they were, rightly, appalled when they were brought to light. This instance of institutional bias only came to light because one employee took it upon himself to act as an agent of change to investigate and correct the problem.

During a training at a large financial services firm, attendees identified some ways that their new employee intake procedures created disparities, offering new employees from "elite" universities significant advantages over those from less "elite" universities. At a large law firm,

attendees identified longstanding gender disparities arising from mentoring activities that involved taking junior associates to a gender-segregated spa, which enabled men, but not women, to have more face-time with the senior partners — all of whom were men. A teacher in one of our trainings realized in-the-moment that, for years, she had been making stereotypic assumptions that disadvantaged her Black and Latin students, whereas she gave her White students the benefit of the doubt. These and many other examples supplement the systemic, empirical research on the bias habit-breaking training to instill confidence that empowerment-based approaches can have widespread, positive effects within individuals and institutions.

I regularly receive emails from attendees, sometimes months or even years after they received the training, sharing how they applied the tools they learned in a way that had a major impact in their lives, sometimes even leading to large-scale changes that address some forms of institutional biases in their organization (e.g., the mail sorting example, above). The training often helps people make concerns more concrete that have previously been vague or uncertain — content from the training gives employees and employers a common language with which to discuss issues related to bias or diversity, and provides solutions for how to address those issues. For me, these anecdotal examples further drive home the importance of a DEI training being equipped to address the infinity variability of bias, as I discussed previously. As someone designing and delivering a bias training, I could never predict or anticipate all the various forms of personal and systemic biases that might be at play in someone's life. However, if a DEI training empowers people as autonomous agents of change who look out for potential biases and are equipped with effective tools to address those biases, most people will make the most of what they learned and do the work to create a more inclusive, less biased world.

### **General Discussion**

I would like to acknowledge that, although they are non-scientific, and often adopt information deficit model approaches, DEI trainings out in the world are likely developed by people who are earnestly invested in having a positive impact, and they likely draw people's attention to factual issues related to bias and diversity. But good intentions alone do not translate to effective approaches to creating change. We all interact with other humans on a daily basis, but that does not make us experts on human behavior any more than the blood in our arteries makes us experts on blood flow dynamics. Human behavior, like blood flow, is governed by underlying lawful processes that extend beyond simple observations. The goal of understanding, predicting, and changing human behavior is best served by the scientific method, and developing DEI trainings is no exception. If one believes in the scientific method and wants to influence human behavior, then one should look to the science of cognitive and behavioral change, and insist on evidence-based, tested approaches to DEI training.

The abundant, increasingly publicized failures of the DEI industry could lead people to feel discouraged, helpless, and defeated with regard to making positive changes related to bias, diversity, equity, and inclusion. Empowerment-based approaches, however, show considerable promise, and give reasons to hope that we can make positive changes. The bias habit-breaking training is just one initial example of the benefit of adopting an empowerment-based approach and believing in people. It has been successful, where so many other trainings have failed, because 1) the training promotes active, self-sustaining change efforts, 2) it teaches customizable, generalizable tools that equip people to address many various forms of bias, 3) it is built on a solid, scientific model of cognitive-behavioral change, and 4) rather than trying to impose change

on people, it respects their autonomy and empowers them to become agents of change themselves. In short, its approach believes the best in people, and helps them to be as effective as possible at living up to their own best intentions. Whether related to DEI or any other domain of human behavior, when we want to create lasting changes, we should believe in science, believe in people, and insist on evidence-based approaches to cognitive-behavioral change.

## References

- al-Gharbi, M. (2020). Diversity-Related Training: What Is It Good For? *Heterodox: The Blog*. Retrieved from: <https://heterodoxacademy.org/blog/diversity-related-training-what-is-it-good-for/> on 11/20/2021.
- Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.
- 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>.
- Apfelbaum, E. P., Norton, M. I., & Sommers, S. R. (2012). Racial colorblindness: Emergence, practice, and implications. *Current Directions in Psychological Science* 21, 205-209.
- Apfelbaum EP, Sommers SR, Norton MI. (2008). Seeing race and seeming racist? Evaluating strategic colorblindness in social interaction. *Journal of Personality and Social Psychology*; 95:918– 932.
- Arnold, K. A., & Loughlin, C. (2013). Integrating transformational and participative versus directive leadership theories. *Leadership & Organization Development Journal*.
- Baker, T., McKay, I., Morden, D. L., Dunning, K., & Schuster, F. E. (1996). Breakthrough in organization performance: Competitive advantage through employee-centered management. *People and Strategy*, 19(4), 14.
- Bargh, J. A., & Tota, M. E. (1988). Context-dependent automatic processing in depression: Accessibility of negative constructs with regard to self but not others. *Journal of Personality and Social Psychology*, 54, 925–939. doi:10.1037/0022-3514.54.6.925
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., & Singmann, H. (2015). lme4: Linear mixed-effects models using Eigen and S4, 2014. R package version, 1(4).
- Beck, A. T. (1967). *Depression: clinical, experimental, and theoretical aspects*. New York, NY: Harper & Row.
- Beck, A. T. (1999). *Prisoners of hate: The cognitive basis of anger, hostility, and violence*. New York, NY: Harper Collins.
- Beck, A. T. (2005). The current state of cognitive therapy: A 40-year retrospective. *Archives of General Psychiatry*, 62, 953–959. doi:10.1001/archpsyc.62.9.953
- Beck, J. S. (2021). *Cognitive therapy: Basics and beyond. Third Edition*. (No. Sirsi) i9780898628470). New York: Guilford press.
- Beck, A. T., & Alford, B. A. (2009). *Depression: Causes and treatment* (2nd ed.). Baltimore, MD: University of Pennsylvania Press.
- Bezrukova K, Spell CS, Perry JL, Jehn KA. 2016. A meta-analytical integration of over 40 years of research on diversity training evaluation. *Psychological Bulletin*. 142(11):1227–74
- Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science*, 16, 162–166. doi:10.1111/j.1467-8721.2007.00496.x
- Birtel, M. D., & Crisp, R. J. (2015). Psychotherapy and social change: Utilizing principles of cognitive-behavioral therapy to help develop new prejudice-reduction interventions. *Frontiers in psychology*, 6, 1771.



- Bodenhausen, G. V., & Macrae, C. N. (1998). Stereotype activation and inhibition. *Advances in social cognition*, 11, 1-52.
- Boykoff, M. (2011). *Who Speaks for the Climate?: Making Sense of Media Reporting on Climate Change*. Cambridge, UK: Cambridge University Press.
- Boynton, P. M. (1998). People should participate in, not be subjects of, research. *BMJ*, 317(7171), 1521.
- Brady, L. M., Kaiser, C. R., Major, B., & Kirby, T. A. (2015). It's fair for us: Diversity structures cause women to legitimize discrimination. *Journal of Experimental Social Psychology*, 57, 100-110.
- Brannon, T. N., Carter, E. R., Murdock-Perriera, L., Higginbotham, G. D. (2018). From backlash to inclusion for all: instituting diversity efforts to maximize benefits across group lines. *Social Issues Policy Review*. 12(1): 57-90
- Brehm, S. S., & Brehm, J. W. (2013). *Psychological reactance: A theory of freedom and control*. Academic Press.
- Brewer, M. B., Hoppel, W. V., & Gooden, M. P. (1999). Diversity and organizational identity: the problem of entrée after entry. In *Cultural Divides: Understanding and Overcoming Group Conflict*, ed. Prentice DA, Miller DT, pp. 337-63. New York, NY: Sage.
- 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.
- Byrnes, D. A., & Kiger, G. (1990). The effect of a prejudice-reduction simulation on attitude change. *Journal of Applied Social Psychology*, 20(4), 341-356.
- Carnes, M. L., Devine, P. G., Manwell, L. B., Byars-Winston, A., Fine, E., Ford, C. E., Forscher, P. S., Iaasc, C., Kaatz, A., Magua, W., Palta, M., & Sherridan, J. (2015). Effect of an intervention to break the gender bias habit: A cluster randomized, controlled trial. *Academic Medicine*, 90
- Carroll, K. M., Ball, S. A., Martino, S., Nich, C., Babuscio, T. A., Nuro, K. F., Gordon, M. A., Portnay, G. A., & Rounsaville, B. J. (2008). Computer-assisted delivery of cognitive-behavioral therapy for addiction: A randomized trial of CBT4CBT. *American Journal of Psychiatry*, 165(7), 881-888.
- Carter, E. R., Onyeador, I. N., & Lewis Jr, N. A. (2020). Developing & delivering effective anti-bias training: Challenges & recommendations. *Behavioral Science & Policy*, 6(1), 57-70.
- Cavanagh, K., Shapiro, D. A., Van Den Berg, S., Swain, S., Barkham, M., & Proudfoot, J. (2006). The effectiveness of computerized cognitive behavioural therapy in routine care. *British Journal of Clinical Psychology*, 45(4), 499-514.
- Chapman, L. J. (1967). Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior*, 6(1), 151-155.
- Cook, D. (2001). Patient autonomy versus paternalism. *Critical Care Medicine*, 29(2), 24-25.
- Cooley, E., Brown-Iannuzzi, J. L., Lei, R. F., & Cipolli, W. III. (2019). Complex intersections of race and class: Among social liberals, learning about White privilege reduces sympathy, increases blame, and decreases external attributions for White people struggling with poverty. *Journal of Experimental Psychology: General*, 148(12), 2218-2228. <https://doi.org/10.1037/xge0000605>

- Cox, W. T. (2015). *Multiple determinants of prejudicial and nonprejudicial behavior* (Doctoral dissertation, The University of Wisconsin-Madison).
- Cox, W. T. L. & Devine, P. G. (2015). Stereotypes possess heterogeneous directionality: A theoretical and empirical exploration of stereotype structure and content. *PLoS ONE* 10(3): e0122292.
- Cox, W. T. L. & Devine, P. G. (2019). The prejudice habit-breaking intervention: An empowerment-based confrontation approach. In Mallett, R. K., & Monteith, M. J. (Eds.). (2019). *Confronting prejudice and discrimination: The science of changing minds and behaviors*. Academic Press., London, UK. 249–274.
- Cox, W. T. L. & Devine, P. G. (2022). Changing implicit bias vs Empowering people to address the personal dilemma of unintentional bias. In *NSF book on Implicit Bias*.
- 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(5), 427-449.
- Cox, W. T. L., Dix, E.L., Scott, D.E., Xie, X., Kellett, K., & Devine, P. G. (2022). Empowering People to Break the Bias Habit: A Randomized-Controlled Field Test of the Updated Bias Habit-Breaking Training. *Preprint available at [www.biashabit.com/research](http://www.biashabit.com/research) after January 1, 2022*.
- Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M., & Huibers, M. J. (2016). How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence. *World Psychiatry*, 15(3), 245-258.
- 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, 784–803.
- Darley J. M., & Gross P. H. (2004). A hypothesis-confirming bias in labeling effects. In *Social Cognition*. Psychology Press, 438-450.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5–18.
- Devine, P. G., & Ash, T. L. (2021). Diversity Training Goals, Limitations, and Promise: A Review of the Multidisciplinary Literature. *Annual review of psychology*, 73.
- Devine, P. G., Forscher, P. S., Austin, A. T., & Cox, W. T. L. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, 4.
- Devine, P.G., Forscher, P.S., Cox, W. T. L., Sherridan, J. Kaatz, A., Carnes, M.L. (2017). A gender habit-breaking intervention led to increased hiring of female faculty in STEMM departments. *Journal of Experimental Social Psychology*, 73, 211-215.
- 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.
- Dickson, D. (2000). Science and its Public: The Need for a Third Way'. *Social studies of science*, 30(6), 917-923.
- Dickson, D. (2005, June 27). *The case for a 'deficit model' of science communication*. Retrieved from <http://www.scidev.net/global/communication/editorials/the-case-for-a-deficit-model-of-science-communic.html>

- Dubé, E., Gagnon, D., & MacDonald, N. E. (2015). Strategies intended to address vaccine hesitancy: Review of published reviews. *Vaccine*, 33(34), 4191-4203.
- Dubé, E., Laberge, C., Guay, M., Bramadat, P., Roy, R., & Bettinger, J. A. (2013). Vaccine hesitancy: an overview. *Human vaccines & immunotherapeutics*, 9(8), 1763-1773.
- 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.
- Dobbin, F., Kalev, A., & Kelly, E. (2007). "Diversity Management in Corporate America." *Contexts* 6(4): 21-7.
- Dobbin, F., Schrage, D., & Kalev, A. (2015). "Rage against the Iron Cage: The Varied Effects of Bureaucratic Personnel Reforms on Diversity." *American Sociological Review* 80(5): 1014-44.
- Dover, T., Major, B., & Kaiser, C. (2014). Diversity initiatives, status, and system-justifying beliefs: When and how diversity efforts de-legitimize discrimination claims. *Group Processes & Intergroup Relations* 17(4): 485-93.
- Duguid, M. & Thomas-Hunt, M. (2015). Condoning Stereotyping? How Awareness of Stereotyping Prevalence Impacts Expression of Stereotypes. *Journal of Applied Psychology* 100(2): 343-59.
- Dunn, E. W., & Spellman, B. A. (2003). Forgetting by remembering: Stereotype inhibition through rehearsal of alternative aspects of identity. *Journal of Experimental Social Psychology*, 39, 420-433. doi:10.1016/S0022-1031(03)00032-5
- Eaves, G., & Rush, A. (1984). Cognitive patterns in symptomatic and remitted unipolar major depression. *Journal of Abnormal Psychology*, 93, 31-40. doi:10.1037/0021843X.93.1.31
- Ebert, D. D., Zarski, A. C., Christensen, H., Stikkelbroek, Y., Cuijpers, P., Berking, M., & Riper, H. (2015). Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: A meta-analysis of randomized controlled outcome trials. *PLOS ONE*, 10(3), e0119895.
- Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & L. Gardner (Eds.), *The handbook of social psychology* (pp. 357-411). New York, NY: McGraw-Hill.
- FitzGerald, C., Martin, A., Berner, D., & Hurst, S. (2019). Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: A systematic review. *BMC Psychology*, 7(1), 29.
- Frisby, C. & O'Donohue, W. (2018). *Cultural Competence in Applied Psychology: An Evaluation of Current Status and Future Directions*. Cham, CH: Springer.
- Folz, C. (2016). "No Evidence That Training Prevents Harassment, Finds EEOC Task Force." *Society for Human Resource Management*, 19 June.
- Forscher P. S. (2017). The individually-targeted habit-breaking intervention and group-level change. [Thesiscommons.org/4t7fy](https://thesiscommons.org/4t7fy)

- Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2019). A meta-analysis of procedures to change implicit measures. *Journal of personality and social psychology*, 117(3), 522.
- Forscher, P. S., Mitamura, C., Cox, W. T. L., Dix, E.L., & Devine, P. G. (2017) Breaking the prejudice habit: Mechanisms, timecourse, and longevity. *J of Experimental Social Psychology*, 72.
- Gibson B. R., Rogers T. T, & Zhu X. (2013). Human semi-supervised learning. *V*, 5(1), 132-72. <https://doi.org/10.1111/tops.12010>
- Green, T. L., & Hagiwara, N. (2020). The Problem with Implicit Bias Training, *Scientific American*
- Hamilton, D. L., & Trolie, T. (1986). Stereotypes and stereotyping: An overview of the cognitive approach. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 127–163). San Diego, CA: Academic Press.
- 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>.
- Hilton, J. L., & von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology*, 47, 237–271. doi:10.1146/annurev.psych.47.1.237
- Himle, J. A., Fischer, D. J., Muroff, J. R., Van Etten, M. L., Lokers, L. M., Abelson, J. L., & Hanna, G. L. (2006). Videoconferencing-based cognitive-behavioral therapy for obsessive-compulsive disorder. *Behaviour Research and Therapy*, 44(12), 1821-1829.
- Hofmann, S. G., Asnaani, A., Vonk, I. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research*, 36(5), 427-440.
- Hofmann, S. G., Asmundson, G. J., & Beck, A. T. (2013). The science of cognitive therapy. *Behavior Therapy*, 44(2), 199-212.
- Hollon, S. D. (2003). Does cognitive therapy have an enduring effect? *Cognitive Therapy and Research*, 27, 71–75. doi:10.1023/A:1022538713914
- Hollon, S. D., & Beck, A. T. (2013). Cognitive and cognitive-behavioral therapies. *Bergin and Garfield's handbook of psychotherapy and behavior change*, 6, 393-442.
- Hollon, S. D., & Dimidjian, S. (2009). Cognitive and behavioral treatment of depression. In I. H. Gotlib & C. L. Hammen (Eds.), *Handbook of depression* (2nd ed., pp. 586–603). New York, NY: Guilford Press.
- Hollon, S. D., & Shelton, R. C. (2001). Treatment guidelines for major depressive disorder. *Behavior Therapy*, 32, 235–258. doi:10.1016/S0005-7894(01)80004-6
- Hoogervorst, J. A. P., Koopman, P. L., & Van der Flier, H. (2005). Total quality management: the need for an employee-centred, coherent approach. *The TQM Magazine*.
- Hornsey, M. J., Harris, E. A., & Fielding, K. S. (2018). The psychological roots of anti-vaccination attitudes: A 24-nation investigation. *Health psychology*, 37(4), 307.
- Howe, L. C., Carr, P. B., & Walton, G. M. (2022). Normative appeals motivate people to contribute to collective action problems more when they invite people to work together toward a

- common goal. *Journal of Personality and Social Psychology*, 121(2), 215–238. <https://doi.org/10.1037/pspa0000278>
- Irwin, A. & Wynne, B. (1996) *Misunderstanding Science? The Public Reconstruction of Science and Technology*. Cambridge University Press.
- Jarrett, C., Wilson, R., O’Leary, M., Eckersberger, E., & Larson, H. J. (2015). Strategies for addressing vaccine hesitancy—A systematic review. *Vaccine*, 33(34), 4180-4190.
- Jensen, J. D., Pokharel, M., Carcioppolo, N., Upshaw, S., John, K. K., & Katz, R. A. (2020). Cancer information overload: Discriminant validity and relationship to sun safe behaviors. *Patient education and counseling*, 103(2), 309-314.
- Kaiser, C. R., Major, B., Jurcevic, I., Dover, T. L., Brady, L. M., & Shapiro, J. R. (2013). Presumed fair: ironic effects of organizational diversity structures. *Journal of personality and social psychology*, 104(3), 504–519. <https://doi.org/10.1037/a0030838>
- Kalish C. W, Rogers T. T, Lang J, & Zhu X. (2011). Can semi-supervised learning explain incorrect beliefs about categories? *Cognition*, 120(1), 106-18.
- 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>.
- Kantor, J. (2020). Behavioral epidemiology: vaccine hesitancy, sunscreen hesitancy, and smoking cessation. *JAAD International*, 1(2), 222-223.
- Kawakami, K., Phillips, C. E., Steele, J. R., & Dovidio, J. F. (2007). (Close) distance makes the heart grow fonder: Improving implicit racial attitudes and interracial interactions through approach behaviors. *Journal of personality and social psychology*, 92(6), 957.
- Kearnes M., Macnaghten P. & Wilsdon, J. (2006). Governing at the Nanoscale. *Demos*; available at <<http://www.demos.co.uk/publications/governingatthenanoscale>>.
- King, M. (2020). Unconscious Bias Training Does Not Work, Here’s How To Fix It. *Forbes*.
- Kulick, C., Perry, E., & Bourhis, A. (2000). Ironic evaluation processes: effects of thought suppression on evaluations of older job applicants. *Journal of Organizational Behaviour* 21(6): 689–711.
- Kumar, R., & Chattu, V. K. (2018). What is in the name? Understanding terminologies of patient-centered, person-centered, and patient-directed care!. *Journal of family medicine and primary care*, 7(3), 487.
- 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(8), 1001–1016.
- Legault, L., Gutsell, J. N., & Inzlicht, M. (2011). Ironic effects of antiprejudice messages how motivational interventions can reduce (but also increase) prejudice. *Psychological Science*.
- Levy, S. R., & Killen, M. (Eds.). (2008). *Intergroup attitudes and relations in childhood through adulthood*. New York, NY: Oxford University Press.
- Lilienfeld, S. O. (2017). Microaggressions: Strong claims, inadequate evidence. *Perspectives on psychological science*, 12(1), 138-169.
- Lipman J. 2018. How diversity training infuriates men and fails women. *Time*.
- Luo, C., Sanger, N., Singhal, N., Patrick, K., Shams, I., Shahid, H., Hoang, P., Schmidt, J., Lee, J., Haber, S., Puckering, M., Buchanan, N., Lee, P., Ng, K., Sun, S., Kheyson, S., Cho-Yan

- Chung, D., Sanger, S., Thabane, L., & Samaan, Z. (2020). A comparison of electronically-delivered and face to face cognitive behavioural therapies in depressive disorders: A systematic review and meta-analysis. *EClinicalMedicine*, 24, 100442.
- 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.
- Magley, V. J., Fitzgerald, L. F., Salisbury, J., Drasgow, F. R. I. T. Z., & Zickar, M. J. (2016). Changing sexual harassment within organizations via training interventions: Suggestions and empirical data. In *The Fulfilling Workplace* (pp. 245-266). Routledge.
- Marteau, T. M., Sowden, A. J., & Armstrong, D. (1998). Implementing research findings into practice: beyond the information deficit model. *Getting research findings into practice*, 36-42.
- McCarthy, W. J. (1985). The cognitive developmental model and other alternatives to the social skills deficit model of smoking onset. *Prevention research: Deterring drug abuse among children and adolescents*, 153.
- McDivitt, P. (2016). The Information Deficit Model is Dead. Now What? Evaluating New Strategies for Communicating Anthropogenic Climate Change in the Context of Contemporary American Politics, Economy, and Culture (Doctoral dissertation, University of Colorado at Boulder).
- McIntosh, P. (1988). *White privilege and male privilege: A personal account of coming to see correspondences through work in women's studies*. Retrieved 14 December, 2005 from the World Wide Web: <http://seamoney.ed.asu.edu/~mei-saac/emc598gunpacking.html>
- Miller, S. (2001). Public understanding of science at the crossroads. *Public Understanding of Science*, 10(1), 115-120. doi:10.1088/0963-6625/10/1/308
- 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.
- 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.
- 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(6171), 615-616.
- Naff, K. & Kellough, J. E. (2007). "Ensuring Employment Equity: Are Federal Diversity Programs Making a Difference?" *International Journal of Public Administration* 26(12): 1307-36.
- Newkirk, P. (2019). *Diversity Inc.: The Failed Promise of a Billion-Dollar Business*. New York, NY: Bold Type Books.
- Nickerson R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2), 175-220. <https://doi.org/10.1037/1089-2680.2.2.175>
- Nordell, J. (2017, May 07). *Is this how discrimination ends?* Retrieved from <https://www.theatlantic.com/science/archive/2017/05/unconscious-bias-training/525405/>.

- Nordell, J. (2021). *The End of Bias: A Beginning: The Science and Practice of Overcoming Unconscious Bias*. Metropolitan Books.
- Norton MI, Sommers SR, Apfelbaum EP, Pura N, Ariely D. (2006). Color Blindness and Interracial Interaction: Playing the Political Correctness Game. *Psychological Science*; 17:949–953.
- Odumeru, J. A., & Ogbonna, I. G. (2013). Transformational vs. transactional leadership theories: Evidence in literature. *International review of management and business research*, 2(2), 355.
- 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.
- 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., Porat, R, Clark, C. S., Green, D. P. (2021). Prejudice reduction: progress and challenges. *Annual Review of Psychology*. 72:14.1-14.28
- Pauker, K., Ambady, N., & Apfelbaum, E. P. (2010). Race salience and essentialist thinking in racial stereotype development: Racial stereotype development. *Child Development*, 81, 1799–1813. doi:10.1111/j.1467-8624.2010.01511.x
- 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>.
- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75, 811-832.
- 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.
- Pohl R. F. (2004). *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory*. Psychology Press.
- Rossen, I., Hurlstone, M. J., & Lawrence, C. (2016). Going with the grain of cognition: applying insights from psychology to build support for childhood vaccination. *Frontiers in Psychology*, 7, 1483.
- Rynes, S. & Rosen, B. (1995). “A Field Survey of Factors Affecting the Adoption and Perceived Success of Diversity Training.” *Personnel Psychology* 48(2): 247-70.
- Saad, Gibbs, Cox, Devine, Gehlbach, McIntyre, & Xie (in progress) Breaking the Bias Habit in Class: Teacher Implicit DEI training Reduces the Achievement Gap
- Sadaf, A., Richards, J. L., Glanz, J., Salmon, D. A., & Omer, S. B. (2013). A systematic review of interventions for reducing parental vaccine refusal and vaccine hesitancy. *Vaccine*, 31(40), 4293–4304.
- Sales, B. D., & Folkman, S. E. (2000). *Ethics in research with human participants*. American Psychological Association.
- Schleider, J., & Weisz, J. (2018). A single-session growth mindset intervention for adolescent anxiety and depression: 9-month outcomes of a randomized trial. *Journal of child psychology and psychiatry, and allied disciplines*, 59(2), 160–170. <https://doi.org/10.1111/jcpp.12811>

- Schultz, P. W. (2002). Knowledge, information, and household recycling: Examining the knowledge-deficit model of behavior change. *New tools for environmental protection: Education, information, and voluntary measures*.
- Siden, J. Y., Carver, A. R., Meja, O. O., Townsel, C. D. (2021). Reducing implicit bias in maternity care: A framework for action. *Women's Health Issues*.
- Simis, M. J., Madden, H., Cacciatore, M. A., & Yeo, S. K. (2016). The lure of rationality: Why does the deficit model persist in science communication?. *Public understanding of science*, 25(4), 400-414.
- Smith, T. C. (2017, July). Vaccine rejection and hesitancy: a review and call to action. In *Open forum infectious diseases* (Vol. 4, No. 3). Oxford University Press.
- Stephan, W. G., & Stephan, C. W. (2001). *Improving intergroup relations*. Thousand Oaks, CA: Sage.
- Stewart, T. L., La Duke, J. R., Bracht, C., Sweet, B. A. M., & Gamarel, K. E. (2003). Do the eyes have it? A program evaluation of Jane Elliott's Blue eyes/Brown eyes diversity training exercise. *Journal of Applied Social Psychology*, 33, 1898–1921.
- Suldozsky, B. (2017). The information deficit model and climate change communication. In *Oxford Research Encyclopedia of Climate Science*.
- Tang, A. & Huang, M. Y. (2021). The problem with implicit bias trainings. *Business Insider*.
- Tønnesen, P. (2002). How to reduce smoking among teenagers. *European Respiratory Journal*, (19).
- Trevors, G., & Duffy, M. C. (2020). Correcting COVID-19 misconceptions requires caution. *Educational Researcher*, 49(7), 538-542.
- 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(2), 207-223.
- Wenzlaff, R. M., Wegner, D. M., & Roper, D. W. (1988). Depression and mental control: The resurgence of unwanted negative thoughts. *Journal of Personality and Social Psychology*, 55, 882– 892. doi:10.1037/0022-3514.55.6.882
- Wilton, L., Apfelbaum, E. & Good, J. (2019). Valuing Differences and Reinforcing Them: Multiculturalism Increases Race Essentialism. *Social Psychological and Personality Science* 10(5): 681-9.
- Wood, W. (2017). Habit in personality and social psychology. *Personality and social psychology review*, 21(4), 389-403.
- Wood, W., & Neal, D. T. (2016). Healthy through habit: Interventions for initiating & maintaining health behavior change. *Behavioral Science & Policy*, 2(1), 71-83.
- Wynne, B. (1992). Misunderstood misunderstanding: social identities and public uptake of science. *Public understanding of science*, 1(3), 281.