Developing scientifically validated bias and diversity trainings that work: empowering agents of change to reduce bias, create inclusion, and promote equity

William Taylor Laimaka Cox

Department of Psychology, University of Wisconsin–Madison, Madison, Wisconsin, USA and Inequity Agents of Change, Madison, Wisconsin, USA

Abstract

Purpose – Research consistently shows that non-scientific bias, equity, and diversity trainings do not work, and often make bias and diversity problems worse. Despite these widespread failures, there is considerable reason for hope that effective, meaningful DEI efforts can be developed. One approach in particular, the bias habit-breaking training, has 15 years of experimental evidence demonstrating its widespread effectiveness and efficacy.

Design/methodology/approach – This article discusses bias, diversity, equity, and inclusion (DEI) efforts from the author's perspective as a scientist–practitioner – the author draws primarily on the scientific literature, but also integrates insights from practical experiences working in DEI. The author provides a roadmap for adapting effective, evidence-based approaches from other disciplines (e.g. cognitive-behavioral therapy) into the DEI context and review evidence related to the bias habit-breaking training, as one prominent demonstration of a scientifically-validated approach that effects lasting, meaningful improvements on DEI issues within both individuals and institutions.

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, in contrast, are highly promising for making meaningful, lasting changes in the DEI realm. Evidence indicates that the bias habit-breaking training is effective at empowering 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 hope and optimism, and an empirically-validated path forward, to develop and test DEI approaches that empower individuals as agents of change.

Keywords Training, Diversity, Equity, Bias, Unconscious bias, DEI

Paper type Research paper

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; Cox and Devine, 2019; Devine and 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 (e.g. Green and Hagiwara, 2020; King, 2020; Lipman, 2018; Tang and 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 content areas outside of DEI. 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, and there are many alternative approaches that more effectively create change (e.g. McDivitt, 2016; Suldovsky, 2017). 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 (Cox and Devine, 2019). 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-breaking training has shown considerable promise at creating lasting, meaningful changes, at both individual and institutional levels (e.g. Cox and Devine, 2019; Cox *et al.*, 2022; Devine *et al.*, 2012, 2017; Forscher *et al.*, 2017).

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 superb work of other authors here, I will briefly state that the overwhelming consensus is that (1) DEI trainings 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; Cox and Devine, 2019; Devine and Ash, 2021; Paluck *et al.*, 2021; Pendry *et al.*, 2007; see also Bezrukova *et al.*, 2016; Brady *et al.*, 2015; Brannon *et al.*, 2018; Brewer *et al.*, 1999; Cooley *et al.*, 2019; Dobbin and Kalev, 2013, 2016; Dobbin *et al.*, 2007, 2015; Dover *et al.*, 2014; Duguid and Thomas-Hunt, 2015; FitzGerald *et al.*, 2019; Frisby and O'Donohue, 2018; Kalev *et al.*, 2006; Kaiser *et al.*, 2013; Legault *et al.*, 2011; Naff and Kellough, 2007; Nordell, 2021; 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 and Devine, 2019; Dobbin and 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). DEI training is estimated to be an \$8 billion industry (Lipman, 2018), and some have gone so far as to declare it a *failed* industry (Newkirk, 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 and 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 deficit 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.

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. Communicators in or from academia may be especially prone to adopting these approaches, because a core tenet of the academic world is to generate, and then share, knowledge. 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; sunscreen use, Jensen *et al.*, 2020; Kantor, 2020; 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 and Duffy, 2020). Despite widespread criticism, abundant evidence of its failure, and even being officially discredited by scientific bodies, the deficit model repeatedly resurfaces (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 echoed by DEI trainings. Efforts in the DEI context can benefit from

Empowering
agents of
change

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 and 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; Sweeny *et al.*, 2010; Wynne, 1992). Developing an intervention that effects meaningful, lasting changes in people's behavior related to a given content area requires not only expertise in that content area, but expertise in intervention science and the science of cognitive-behavioral change.

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 activities (e.g. Carter *et al.*, 2020; Devine and Ash, 2021; Paluck and Green, 2009; Paluck *et al.*, 2021; Pendry *et al.*, 2007). As reviewed extensively by Pendry *et al.* (2007), very common DEI training content involves educating people about historical intergroup disparities, the existence of biases/disparities in modern society or the workplace, implicit bias, White privilege (McIntosh, 1988), microaggressions (Lilienfeld, 2017), how it feels to be excluded/ discriminated against (Byrnes and Kiger, 1990; Stewart *et al.*, 2003), or the benefits of diverse workforces. 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 their own change process (see Cox and Devine, 2019). Rather than change being something that is done *to* people, empowerment-based approaches focus on working *with* people, and empowering them as autonomous 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 and 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 and 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 and Chattu, 2018). Similar discussions in management science have arisen in recent years, emphasizing transformational leadership over transactional and directive leadership styles (Arnold and Loughlin, 2013; Odumeru and 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 emphasized the benefits of moving away from paternalistic approaches 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. Fortuitously, basic research has long demonstrated considerable reasons for optimism about trusting 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 and 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 intergroup biases (Amodio *et al.*, 2007; Cox, 2015; Devine, 1989; Devine *et al.*, 1991; Monteith, 1993; Monteith *et al.*, 2001, 2002; Plant and Devine, 1998, 2009) [1].

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, 2012; Kulick *et al.*, 2000; Norton *et al.*, 2006; Uhlmann and Cohen, 2007). At the individual level, for example, trying to suppress stereotypes leads a rebound effect where people display *more* bias, rather than less (Macrae *et al.*, 1994). At an institutional level, implementing trainings that use the information deficit model, as discussed above, likewise backfire. 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 identify 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 and Macrae, 1998; Cox and 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 and Liben, 2007; Levy and 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; Cox *et al.*, 2022; Darley and 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 and Neal, 2016) and to approach bias reduction as a process of breaking a habit (Cox and 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 (e.g. Cox *et al.*, 2022; Reggev *et al.*, 2021). 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 and 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 and Tota, 1988; Beck, 1967; Beck and Alford, 2009; Devine, 1989; Dunn and Spellman, 2003; Eaves and Rush, 1984; Fiske, 1998; Hamilton and Trolier, 1986; Hilton and von Hippel, 1996; Wenzlaff *et al.*, 1988). Building on this insight, Cox *et al.* (2012) argued that methods from experimental clinical research, especially cognitive-behavioral therapy (CBT), could be useful for DEI efforts (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, teaching them to understand the impact of those 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, 2021; Hollon, 2003; Hollon and Dimidjian, 2009; Hollon and 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, 2021; Hollon, 2003; Hollon and Beck, 2013; Hollon and Dimidjian, 2009; Hollon and 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 and 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). CBT is also a very adaptable framework that can incorporate aspects of other approaches (e.g. mindfulness therapy, motivational interviewing) as needed or as appropriate. 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 *et al.* (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 for DEI efforts, 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, age, many other social identities, and the intersections thereof (Cox *et al.*, 2012; Crenshaw, 1989). It is infeasible for organizations to implement separate interventions or training programs for every stereotyped group that may be affected by biases/disparities (e.g. Black people, Latin people, Asian people, Muslim people, women, LGBT + people, people with disabilities, and their intersections). 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 forms they take, biases and inequities have truly limitless variability. 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, teach skills that enable recipients 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 must likewise teach generalizable and customizable skills, so recipients can 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 method to address the infinite varieties of bias and inequity.

The bias habit-breaking training: empowering agents of change

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; Cox *et al.*, 2022; Devine *et al.*, 2012; for another review see Cox and 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 bias habit-breaking 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 *et al.* (2012) framework, the latest versions of the bias habit-breaking training adapt core principles from CBT to help meet these goals (Beck, 2021; see Table 1).

CBT…	The bias habit-breaking training…
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 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 disparities and inequity, and teaches them to disrupt those biases
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

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

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, and to self-sustain this process over time. 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 biases and inequity are discussed in a very evidence-based, non-accusatory way – it frames the scientific evidence as demonstrating people's unwitting *vulnerability* to biases, rather than presenting bias as inevitable or indicative of someone's moral character. 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 and efficacy

The present article has drawn on both my expertise as a cognitive-behavioral scientist and my applied experience as a DEI practitioner. As a scientist, I rely on systematic, experimental data 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. I draw on both kinds of evidence in the following sections: In addition to reviewing published and ongoing quantitative, empirical research testing the bias habit-breaking training, I will also share a few (admittedly anecdotal) examples of real-world impacts the training has had on individuals and institutions.

Several experimental studies have demonstrated lasting and impactful effects of the bias habit-breaking training. As noted earlier, the training has been iteratively tested, 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).

Following several randomized-controlled experiments demonstrating the effectiveness of the bias habit-breaking training, my colleagues and I began receiving requests to deliver the training live, independent of any formal empirical study. Given the evidence we had amassed at that point, we felt confident transitioning from testing to practical application, and we have delivered the training live for many companies, organizations, academic institutions, and governmental agencies around the world. Through these practical endeavors, we have been able to witness firsthand how the bias habit-breaking training impacts people and organizations. 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 or contributing to inequity without realizing it. Very often, these are highly specific forms of bias or inequity that are particular to the individual or a particular workplace – examples unlikely to be captured in typical empirical, quantitative studies. In the following sections, after reviewing the empirical findings, I share a selection of these firsthand examples, to showcase the breadth of potential outcomes that can be influenced when individuals are situated and empowered as the primary agents of change.

I organize the review of these empirical findings and firsthand examples under three broad themes that correspond to key areas where DEI programs seek to make progress: Reducing Bias, Creating Inclusion, and Promoting Equity.

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 for discussion of these constructs' importance). 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 race bias 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 remains so today. In a recent meta-analysis of 492 experiments trying to reduce implicit bias (Forscher *et al.*, 2019), no study demonstrated decreases in implicit bias that lasted more than 24 h, with most lasting only a few minutes (see also Lai *et al.*, 2016; Siden *et al.*, 2022). Work on the bias habit-breaking training is the sole exception to these patterns.

In another study, 302 undergraduate participants were randomly assigned to receive the bias habit-breaking training or to serve as controls (Forscher *et al.*, 2017). These participants completed IATs every other day for two weeks, in an attempt to model the time-course of the previously observed reduction in IAT bias. Replicating Devine *et al.* (2012) pattern, the training participants' IAT scores decreased over time. Contrasting with the prior study, however, the control participants' scores also decreased. Given that the participants completed the IAT with such a high frequency (up to 8 times in a two-week period), it seemed that this pattern constituted a practice effect on the IAT task. This interpretation matches other studies that show learning effects in the IAT task (e.g. Cochrane *et al.*, 2022). Devine *et al.* (2012) original pattern of reduced implicit bias only in the training condition 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), participants completed the IAT two weeks and six weeks after random assignment, and 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. This pattern further supports the interpretation that the Forscher *et al.* data reflects a practice effect on the IAT, whereas the other two studies' reductions in the training, but not control, conditions are more likely due to training participants implementing the bias reduction tools taught in the training.

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 and Devine, 2022). In the Forscher *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 *et al.*, 2019). Again, work with the bias habit-breaking training appears to be the sole exception. A subsample ($N = 320$) of Cox *et al.* (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. These participants completed Monteith *et al.* (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 *et al.*, 2019), the reductions resulting from the bias habit-breaking training appear to be both long-lasting and related to other, meaningful outcomes.

Initial experimental tests of the training involved solely or mostly White participants, either by design (Devine *et al.*, 2012; Forscher *et al.*, 2017) or because the sample merely reflected the demographics of the predominantly White study population (Cox *et al.*, 2022). In the interest of focusing on Black voices in this line of research, Auxier and Cox (2022) specifically recruited Black participants to complete the training and to share their perspectives and opinions about it, via an evaluation survey ($N = 28$) immediately after the training, and a qualitative interview with a Black interviewer ($N = 12$) at least two weeks after the training. One key concern was that, due to their frequent experiences as targets of bias, Black participants might see the training as irrelevant to their own cognitions and behavior. Contrary to this expectation, all 12 participants who completed interviews reported that they had applied what they learned in the training, either by directly using one of the bias-reducing tools, or by applying the content in a way that improved interactions with others. All 12 interviewees also reported that the training was respectful and appropriate in its treatments of biases and other issues related to Black people. Mirroring the increase in awareness observed in the initial studies with predominantly White participants, several of the interviewees reported that the training helped them tune in to forms of bias they might themselves express toward other social groups.

These patterns with Auxier and Cox's participants are reflected in many of my firsthand experiences delivering the training. Many attendees who are members of historically disadvantaged groups (e.g. people of color, women, LGBT + people) have reported realizing that they may have disregarded their potential to express biases based on other group statuses, perhaps due to predominant social rhetoric that frames biases as arising solely or primarily from members of historically powerful groups. Because the training discusses bias as arising from ordinary cognitive processes that occur within everyone – not just White people, or men, or straight people, or people from other non-stigmatized social groups – it empowers everyone to look for and notice biases they might be vulnerable to expressing.

Indeed, attendees often realize, either in-the-moment or after the training, various ways they have been expressing biases without realizing it, and then make a plan to stop that bias. A teacher in one of our trainings was mortified when she realized, in-the-moment that, for years, she had been making stereotypic assumptions that disadvantaged her Black and Latin students. Specifically, if a Black or Latin student failed to turn in an assignment or came late to class, she would make the snap judgment that they were lazy and did not care about their coursework. When White students engaged in similar behavior, however, she more often gave them the benefit of the doubt and gave them opportunities to explain themselves. During another training, an attendee received a message from an employee saying he was going to be late to work that day. The attendee began getting angry with the employee, jumping to conclusions about him not caring about his job. This circumstance unfolded during the section of the training in which I explain tools that help people retrain their reactions to curb snap judgments. She applied one of these tools (*considering situational explanations for behavior;* e.g. Stewart *et al.*, 2010) while I was discussing it, and immediately began to feel more calm, as her generated situational explanations for the employee's behavior (e.g. "maybe his car broke down," "maybe his childcare fell through") gave her more compassion for him. She told me about the experience afterward via email and shared that her relationship with that employee had noticeably improved. Many more firsthand examples abound, and they reinforce the idea that some forms of bias are highly common across many people, and some are more specific or unique to a particular person or circumstance. Because

the bias habit-breaking training puts people on a path to identify and correct biases themselves, it is effective at equipping them to identify and address bias in its myriad forms.

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 speak up and confront biased or offensive rhetoric. Testing the training's effects on speaking up, Forscher *et al.* (2017) followed up with their participants via an ostensibly unrelated study 1–2 years post-manipulation. Training and control participants received an email that they believed was sent by their university 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 authors' 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 *et al.* (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. Importantly, these two DEI topics were not mentioned in the content of the training that the training participants had completed prior, thus enabling Cox *et al.* 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 an inclusive social climate 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 (Auxier and Cox, 2022; Cox *et al.*, 2022; Devine *et al.*, 2012; Forscher *et al.*, 2017). In both lab studies and real-world experimental 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 may suggest 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.

Organizational climate is another important indicator of creating inclusion. 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 *et al.* (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).

Several studies on the bias habit-breaking training have shown evidence of creating inclusion in the form of increased or improved intergroup interactions. In free response data, training participants often report seeking out or being open to more interactions with members of other groups, at rates higher than those of control participants (Cox *et al.*, 2022; Devine *et al.*, 2012; Forscher *et al.*, 2017). In Auxier and Cox's (2022) exploration of Black participants' perspectives on the training, 9 of the 12 interviewees reported that the training changed how they interpreted other people's potentially biased behaviors, leading them to be more understanding when someone else expresses bias unintentionally. 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.

An attendee wrote me a year after he attended a session of the bias habit-breaking training to tell me about how the training changed his actions in a way that made a difference for his whole neighborhood. A few months after he attended the training, a Mexican family moved into his predominantly White neighborhood. His initial thought was that he and his family should not go introduce themselves to the new family, because, "They do not want some White people coming to bother them!". But then he remembered part of the training that explicitly warned against making assumptions that people from different backgrounds will lack interest in having interactions (cf. "pluralistic ignorance", Shelton and Richeson, 2005). He overcame his hesitance and took his family to welcome their new neighbors. Not only did this attendee and his family become close friends with their new neighbors, but their inclusive gesture inspired others in the neighborhood to overcome their own apprehensions, and, in this attendee's words, it "set off a chain reaction" of welcoming and outreach.

Promoting equity. Inequity can take many forms, in different contexts. In academic 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 during the two years prior to random assignment revealed that new tenure-track hires in both intervention and control departments were only 32–33% women. In the two years after departments either received the training or were randomly assigned to be controls, new hires in control departments were still only 32% women, but new hires in intervention departments were 47% women, Contrast

OR = 1.89, p = 0.0109, BF_{-0} = 30.40. See Figure 1. In other words, the bias habit-breaking training caused a 43% increase in hiring of members of underrepresented groups.

These general hiring data were previously explored in Devine *et al.* (2017), but these analyses and figure are new for the present publication. The bar graph shows raw population percentages of women hired as new faculty during the 2-year periods before and after random assignment to intervention or control departments, as reported in Carnes *et al.* (2015). We conducted a Generalized Linear Mixed Effects Model set up as reported in Devine *et al.* (2017), except that we tested a 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 in the two years after the training was significantly higher than intervention departments in the two years before the training, or the control departments two years before or two years after the manipulation, OR = 1.89, p = 0.0109, 95% CI = [1.16, 3.11]. A Bayesian Independent Samples *T*-test (BF_{-0} = 30.40) for this contrast further indicated these data constitute *very strong* evidence that the bias training caused increased hiring of women.

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 7,084 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*).

Inequity becomes integrated into social systems and institutions in myriad forms, many of which are unique to specific organizations. This uniqueness makes it crucial that DEI efforts center on individuals within an organization as the agents of change (as in empowerment-based approaches), rather than the trainer as the driver of change (as in information deficit model approaches), because those *in* the day-to-day culture of an organization are the ones
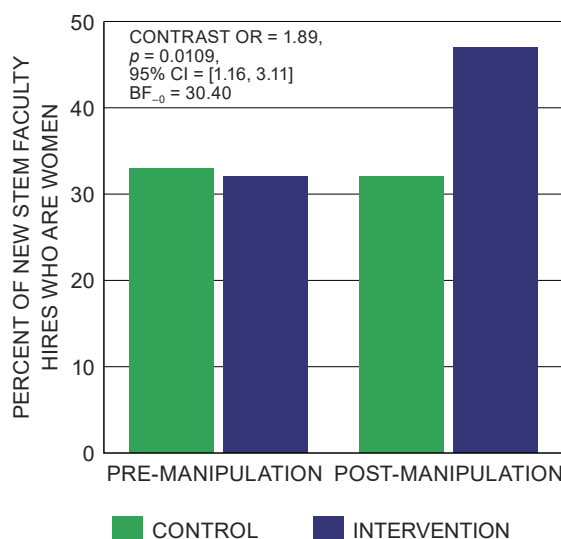
Figure 1.
Hiring data - percent of new tenure-track faculty hires who were women

best equipped to identify and correct their organization's particular inequities. At one company that received the bias habit-breaking training, learning about how inequities persist within organizational structures led an employee to investigate the method used by their company's automatic 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 inequity 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, and immediately changed their onboarding procedure to correct this inequity. 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. 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/inequities 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 often non-scientific and often adopt information deficit model approaches, DEI trainings out in the world are predominantly 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 effectively creating meaningful 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, experimentally 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.

Notes

1. For the minority of people whose personal values consciously *endorse* prejudice, other kinds of approaches (e.g. strong anti-prejudice social norms) might reduce the open expression of bias. It is unlikely, however, that *any* intervention will influence these people who are highly motivated to express prejudice to become positive agents of change (Forscher *et al.*, 2015). Luckily, most people do not fall into this category.

References

al-Gharbi, M. (2020), "Diversity-related training: what is it good for?", *Heterodox: The Blog*, available at: https://heterodoxacademy.org/blog/diversity-related-training-what-is-it-good-for/on 11/20/2021.

Allport, G.W. (1954), *The Nature of Prejudice*, Addison-Wesley, Reading, MA.

Amodio, D.M., Devine, P.G. and Harmon-Jones, E. (2007), "A dynamic model of guilt: implications for motivation and self-regulation in the context of prejudice", *Psychological Science*, Vol. 18, pp. 524-530, doi: 10.1111/j.1467-9280.2007.01933.x.

Apfelbaum, E.P., Sommers, S.R. and Norton, M.I. (2008), "Seeing race and seeming racist? Evaluating strategic colorblindness in social interaction", *Journal of Personality and Social Psychology*, Vol. 95, pp. 918-932.

Apfelbaum, E.P., Norton, M.I. and Sommers, S.R. (2012), "Racial colorblindness: emergence, practice, and implications", *Current Directions in Psychological Science*, Vol. 21, pp. 205-209.

Arnold, K.A. and Loughlin, C. (2013), "Integrating transformational and participative versus directive leadership theories", *Leadership and Organization Development Journal*, Vol. 34 No. 1, pp. 67-84.

Auxier, S.R. and Cox, W.T.L. (2022), "Amplifying Black voices on campus: black students' reactions to the bias habit-breaking training", *Undergraduate Research Symposium, University of Wisconsin-Madison, Madison, WI*.

Baker, T., McKay, I., Morden, D.L., Dunning, K. and Schuster, F.E. (1996), "Breakthrough in organization performance: competitive advantage through employee-centered management", *People and Strategy*, Vol. 19 No. 4, p. 14.

Bargh, J.A. and 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*, Vol. 54, pp. 925-939, doi: 10.1037/0022-3514.54.6.925.

Beck, A.T. (1967), *Depression: Clinical, Experimental, and Theoretical Aspects*, Harper & Row, New York, NY.

Beck, A.T. (1999), *Prisoners of Hate: The Cognitive Basis of Anger, Hostility, and Violence*, Harper Collins, New York, NY.

Beck, A.T. (2005), "The current state of cognitive therapy: a 40-year retrospective", *Archives of General Psychiatry*, Vol. 62, pp. 953-959, doi: 10.1001/archpsyc.62.9.953.

Beck, J.S. (2021), *Cognitive Therapy: Basics and Beyond*, (No. Sirsi) i9780898628470), 3rd ed., Guilford press, New York, NY.

Beck, A.T. and Alford, B.A. (2009), *Depression: Causes and Treatment*, 2nd ed., University of Pennsylvania Press, Baltimore, MD.

Bezrukova, K., Spell, C.S., Perry, J.L. and Jehn, K.A. (2016), "A meta-analytical integration of over 40 years of research on diversity training evaluation", *Psychological Bulletin*, Vol. 142 No. 11, pp. 1227-1274.

Bigler, R.S. and Liben, L.S. (2007), "Developmental intergroup theory: explaining and reducing children's social stereotyping and prejudice", *Current Directions in Psychological Science*, Vol. 16, pp. 162-166, doi: 10.1111/j.1467-8721.2007.00496.x.

Bodenhausen, G.V. and Macrae, C.N. (1998), "Stereotype activation and inhibition", *Advances in Social Cognition*, Vol. 11, pp. 1-52.

Boykoff, M. (2011), *Who Speaks for the Climate?: Making Sense of Media Reporting on Climate Change*, Cambridge University Press, Cambridge.

Boynton, P.M. (1998), "People should participate in, not be subjects of, research", *BMJ*, Vol. 317 No. 7171, p. 1521.

Brady, L.M., Kaiser, C.R., Major, B. and Kirby, T.A. (2015), "It's fair for us: diversity structures cause women to legitimize discrimination", *Journal of Experimental Social Psychology*, Vol. 57, pp. 100-110.

Brannon, T.N., Carter, E.R., Murdock-Perriera, L. and Higginbotham, G.D. (2018), "From backlash to inclusion for all: instituting diversity efforts to maximize benefits across group lines", *Social Issues Policy Review*, Vol. 12 No. 1, pp. 57-90.

Brehm, S.S. and Brehm, J.W. (2013), *Psychological Reactance: A Theory of Freedom and Control*, Academic Press, New York.

Brewer, M.B., Hippel, W.V. and Gooden, M.P. (1999), "Diversity and organizational identity: the problem of entrée after entry", in Prentice, D.A. and Miller, D.T. (Eds), *in Cultural Divides: Understanding and Overcoming Group Conflict*, Sage, New York, NY, pp. 337-363.

Burns, M.D., Monteith, M.J. and 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*, Vol. 73, pp. 97-110.

Byrnes, D.A. and Kiger, G. (1990), "The effect of a prejudice-reduction simulation on attitude change", *Journal of Applied Social Psychology*, Vol. 20 No. 4, pp. 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. and Sherridan, J. (2015), "Effect of an intervention to break the gender bias habit: a cluster randomized, controlled trial", *Academic Medicine*, Vol. 90, pp. 221-230.

Carroll, K.M., Ball, S.A., Martino, S., Nich, C., Babuscio, T.A., Nuro, K.F., Gordon, M.A., Portnay, G.A. and Rounsaville, B.J. (2008), "Computer-assisted delivery of cognitive-behavioral therapy for

addiction: a randomized trial of CBT4CBT", *American Journal of Psychiatry*, Vol. 165 No. 7, pp. 881-888.

Carter, E.R., Onyeador, I.N. and Lewis, N.A., Jr (2020), "Developing and delivering effective anti-bias training: challenges and recommendations", *Behavioral Science and Policy*, Vol. 6 No. 1, pp. 57-70.

Cavanagh, K., Shapiro, D.A., Van Den Berg, S., Swain, S., Barkham, M. and Proudfoot, J. (2006), "The effectiveness of computerized cognitive behavioural therapy in routine care", *British Journal of Clinical Psychology*, Vol. 45 No. 4, pp. 499-514.

Chapman, L.J. (1967), "Illusory correlation in observational report", *Journal of Verbal Learning and Verbal Behavior*, Vol. 6 No. 1, pp. 151-155.

Cochrane, A., Cox, W.T.L. and Green, S. (2022), "The implicit association test is sensitive to learning: exploring within-session modulations of IAT scores", Under review.

Cook, D. (2001), "Patient autonomy versus paternalism", *Critical Care Medicine*, Vol. 29 No. 2, pp. 24-25.

Cooley, E., Brown-Iannuzzi, J.L., Lei, R.F. and 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*, Vol. 148 No. 12, pp. 2218-2228, doi: 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. and Devine, P.G. (2015), "Stereotypes possess heterogeneous directionality: a theoretical and empirical exploration of stereotype structure and content", *PLoS ONE*, Vol. 10 No. 3, e0122292.

Cox, W.T.L. and Devine, P.G. (2019), "The prejudice habit-breaking intervention: an empowerment-based confrontation approach", in Mallett, R.K. and Monteith, M.J. (Eds), *Confronting Prejudice and Discrimination: The Science of Changing Minds and Behaviors*, Academic Press, London, pp. 249-274.

Cox, W.T.L. and Devine, P.G. (2022), "Changing implicit bias vs Empowering people to address the personal dilemma of unintentional bias", *The Cambridge Handbook of Implicit Bias and Racism*, Cambridge University Press, Cambridge.

Cox, W.T.L., Abramson, L.Y., Devine, P.G. and Hollon, S.D. (2012), "Stereotypes, prejudice, and depression: the integrated perspective", *Perspectives on Psychological Science*, Vol. 7 No. 5, pp. 427-449.

Cox, W.T.L., Dix, E.L., Scott, D.E., Xie, X., Kellett, K. and Devine, P.G. (2022), "Empowering people to break the bias habit: a randomized-controlled field test of the updated bias habit-breaking training", available at: www.biashabit.com/research (accessed 1 January 2022).

Cox, W.T.L., Xie, X. and Devine, P.G. (2022), "Untested assumptions perpetuate stereotyping: Learning in the absence of evidence", *Journal of Experimental Social Psychology*, Vol. 102, doi: 10.1016/j.jesp.2022.104380.

Crenshaw, K. (1989), *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory, and Antiracist Politics*, University of Chicago Legal Forum, Chicago, pp. 139-167.

Cuijpers, P., Cristea, I.A., Karyotaki, E., Reijnders, M. and 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*, Vol. 15 No. 3, pp. 245-258.

Czopp, A.M., Monteith, M.J. and Mark, A.Y. (2006), "Standing up for a change: reducing bias through interpersonal confrontation", *Journal of Personality and Social Psychology*, Vol. 90, pp. 784-803.

Darley, J.M. and Gross, P.H. (2004), "A hypothesis-confirming bias in labeling effects", in *Social Cognition*, Psychology Press, pp. 438-450.

Devine, P.G. (1989), "Stereotypes and prejudice: their automatic and controlled components", *Journal of Personality and Social Psychology*, Vol. 56, pp. 5-18.

Devine, P.G. and Ash, T.L. (2021), "Diversity training goals, limitations, and promise: a review of the multidisciplinary literature", *Annual Review of Psychology*, Vol. 73, pp. 403-411.

Devine, P.G., Monteith, M.J., Zuwerink, J.R. and Elliot, A.J. (1991), "Prejudice with and without compunction", *Journal of Personality and Social Psychology*, Vol. 60, pp. 817-830.

Devine, P.G., Forscher, P.S., Austin, A.T. and Cox, W.T.L. (2012), "Long-term reduction in implicit race bias: a prejudice habit-breaking intervention", *Journal of Experimental Social Psychology*, Vol. 4, pp. 1268-1278.

Devine, P.G., Forscher, P.S., Cox, W.T.L., Sherridan, J., Kaatz, A. and Carnes, M.L. (2017), "A gender habit-breaking intervention led to increased hiring of female faculty in STEMM departments", *Journal of Experimental Social Psychology*, Vol. 73, pp. 211-215.

Dickson, D. (2000), "Science and its public: the need for aThird way", *Social Studies of Science*, Vol. 30 No. 6, pp. 917-923.

Dickson, D. (2005), "The case for a 'deficit model' of science communication", available at: http://www.scidev.net/global/communication/editorials/the-case-for-a-deficit-model-of-science-communic.html.

Dobbin, F. and Kalev, A. (2013), "The origins and effects of corporate diversity programs", in Roberson, Q.M. (Ed.), *The Oxford Handbook of Diversity and Work*, Oxford University Press, Oxford, pp. 253-281.

Dobbin, F. and Kalev, A. (2016), "Why diversity programs fail", *Harvard Business Review*, Vol. 94, pp. 52-61.

Dobbin, F., Kalev, A. and Kelly, E. (2007), "Diversity management in corporate America", *Contexts*, Vol. 6 No. 4, pp. 21-27.

Dobbin, F., Schrage, D. and Kalev, A. (2015), "Rage against the iron cage: the varied effects of bureaucratic personnel reforms on diversity", *American Sociological Review*, Vol. 80 No. 5, pp. 1014-1044.

Dover, T., Major, B. and Kaiser, C. (2014), "Diversity initiatives, status, and system-justifying beliefs: when and how diversity efforts de-legitimize discrimination claims", *Group Processes and Intergroup Relations*, Vol. 17 No. 4, pp. 485-493.

Dubé, E., Laberge, C., Guay, M., Bramadat, P., Roy, R. and Bettinger, J.A. (2013), "Vaccine hesitancy: an overview", *Human Vaccines and Immunotherapeutics*, Vol. 9 No. 8, pp. 1763-1773.

Dubé, E., Gagnon, D. and MacDonald, N.E. (2015), "Strategies intended to address vaccine hesitancy: review of published reviews", *Vaccine*, Vol. 33 No. 34, pp. 4191-4203.

Duguid, M. and Thomas-Hunt, M. (2015), "Condoning stereotyping? How awareness of stereotyping prevalence impacts expression of stereotypes", *Journal of Applied Psychology*, Vol. 100 No. 2, pp. 343-359.

Dunn, E.W. and Spellman, B.A. (2003), "Forgetting by remembering: stereotype inhibition through rehearsal of alternative aspects of identity", *Journal of Experimental Social Psychology*, Vol. 39, pp. 420-433, doi: 10.1016/S0022-1031(03)00032-5.

Eaves, G. and Rush, A. (1984), "Cognitive patterns in symptomatic and remitted unipolar major depression", *Journal of Abnormal PsyChology*, Vol. 93, pp. 31-40, doi: 10.1037/0021843X.93.1.31.

Ebert, D.D., Zarski, A.C., Christensen, H., Stikkelbroek, Y., Cuijpers, P., Berking, M. and 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*, Vol. 10 No. 3, e0119895.

Fiske, S.T. (1998), "Stereotyping, prejudice, and discrimination", in Gilbert, D.T., Fiske, S.T. and Gardner, L. (Eds), *The Handbook of Social Psychology*, McGraw-Hill, New York, NY, pp. 357-411.

FitzGerald, C., Martin, A., Berner, D. and Hurst, S. (2019), "Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: a systematic review", *BMC Psychology*, Vol. 7 No. 1, p. 29.

Forscher, P.S. (2017), "The individually-targeted habit-breaking intervention and group-level change", available at: thesiscommons.org/4t7fy.

Forscher, P.S., Cox, W.T., Graetz, N. and Devine, P.G. (2015), "The motivation to express prejudice", *Journal of Personality and Social Psychology*, Vol. 109 No. 5, p. 791.

Forscher, P.S., Mitamura, C., Cox, W.T.L., Dix, E.L. and Devine, P.G. (2017), "Breaking the prejudice habit: mechanisms, timecourse, and longevity", *Journal of Experimental Social Psychology*, Vol. 72, pp. 133-146.

Forscher, P.S., Lai, C.K., Axt, J.R., Ebersole, C.R., Herman, M., Devine, P.G. and Nosek, B.A. (2019), "A meta-analysis of procedures to change implicit measures", *Journal of Personality and Social Psychology*, Vol. 117 No. 3, p. 522.

Frisby, C. and O'Donohue, W. (2018), *Cultural Competence in Applied Psychology: An Evaluation of Current Status and Future Directions*, Springer, Cham.

Gibson, B.R., Rogers, T.T. and Zhu, X. (2013), "Human semi-supervised learning", *Topics in Cognitive Science*, Vol. 5 No. 1, pp. 132-172, doi: 10.1111/tops.12010.

Green, T.L. and Hagiwara, N. (2020), *The Problem with Implicit Bias Training*, Scientific American, available at: https://www.scientificamerican.com/article/the- problem-with-implicit-bias-training/.

Hamilton, D.L. and Trolier, T. (1986), "Stereotypes and stereotyp- ing: an overview of the cognitive approach", in Dovidio, J.F. and Gaertner, S.L. (Eds), *Prejudice, Discrimination, and Racism*, Academic Press, San Diego, CA, pp. 127-163.

Hennes, E.P., Pietri, E.S., Moss-Racusin, C.A., Mason, K.A., Dovidio, J.F., Brescoll, V.L., Bailey, H., A. and Handelsman, J. (2018), "Increasing the perceived malleability of gender bias using a modified video intervention for diversity in STEM (VIDS)", *Group Processes and Intergroup Relations*, Vol. 21 No. 5, pp. 788-809, doi: 10.1177/1368430218755923.

Hilton, J.L. and von Hippel, W. (1996), "Stereotypes", *Annual Review of Psychology*, Vol. 47, pp. 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. and Hanna, G.L. (2006), "Videoconferencing-based cognitive-behavioral therapy for obsessive-compulsive disorder", *Behaviour Research and Therapy*, Vol. 44 No. 12, pp. 1821-1829.

Hofmann, S.G., Asnaani, A., Vonk, I.J., Sawyer, A.T. and Fang, A. (2012), "The efficacy of cognitive behavioral therapy: a review of meta-analyses", *Cognitive Therapy and Research*, Vol. 36 No. 5, pp. 427-440.

Hofmann, S.G., Asmundson, G.J. and Beck, A.T. (2013), "The science of cognitive therapy", *Behavior Therapy*, Vol. 44 No. 2, pp. 199-212.

Hollon, S.D. (2003), "Does cognitive therapy have an enduring effect?", *Cognitive Therapy and Research*, Vol. 27, pp. 71-75, doi: 10.1023/A:1022538713914.

Hollon, S.D. and Beck, A.T. (2013), "Cognitive and cognitive-behavioral therapies", *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change*, Vol. 6, pp. 393-442.

Hollon, S.D. and Dimidjian, S. (2009), "Cognitive and behavioral treatment of depression", in Gotlib, I.H. and Hammen, C.L. (Eds), *Handbook of Depression*, 2nd ed., Guilford Press, New York, NY, pp. 586-603.

Hollon, S.D. and Shelton, R.C. (2001), "Treatment guidelines for major depressive disorder", *Behavior Therapy*, Vol. 32, pp. 235-258, doi: 10.1016/S0005-7894(01)80004-6.

Hoogervorst, J.A.P., Koopman, P.L. and van der Flier, H. (2005), "Total quality management: the need for an employee-centred, coherent approach", *The TQM Magazine*, Vol. 17, pp. 92-106.

Hornsey, M.J., Harris, E.A. and Fielding, K.S. (2018), "The psychological roots of anti-vaccination attitudes: a 24-nation investigation", *Health Psychology*, Vol. 37 No. 4, p. 307.

Howe, L.C., Carr, P.B. and 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*, Vol. 121 No. 2, pp. 215-238, doi: 10.1037/pspa0000278.

Irwin, A. and Wynne, B. (1996), *Misunderstanding Science? the Public Reconstruction of Science and Technology*, Cambridge University Press, Cambridge.

Jarrett, C., Wilson, R., O'Leary, M., Eckersberger, E. and Larson, H.J. (2015), "Strategies for addressing vaccine hesitancy–A systematic review", *Vaccine*, Vol. 33 No. 34, pp. 4180-4190.

Jensen, J.D., Pokharel, M., Carcioppolo, N., Upshaw, S., John, K.K. and Katz, R.A. (2020), "Cancer information overload: discriminant validity and relationship to sun safe behaviors", *Patient Education and Counseling*, Vol. 103 No. 2, pp. 309-314.

Kaiser, C.R., Major, B., Jurcevic, I., Dover, T.L., Brady, L.M. and Shapiro, J.R. (2013), "Presumed fair: ironic effects of organizational diversity structures", *Journal of Personality and Social Psychology*, Vol. 104 No. 3, pp. 504-519, doi: 10.1037/a0030838.

Kalev, A., Dobbin, F. and Kelly, E. (2006), "Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies", *American Sociological Review*, Vol. 71, pp. 589-617, doi: 10.1177/000312240607100404.

Kalish, C.W., Rogers, T.T., Lang, J. and Zhu, X. (2011), "Can semi-supervised learning explain incorrect beliefs about categories?", *Cognition*, Vol. 120 No. 1, pp. 106-118.

Kantor, J. (2020), "Behavioral epidemiology: vaccine hesitancy, sunscreen hesitancy, and smoking cessation", *JAAD International*, Vol. 1 No. 2, pp. 222-223.

Kearnes, M., Macnaghten, P. and 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*, available at: https://www.forbes.com/sites/michelleking/2020/11/10/unconcious-bias-training-does-not-work-heres-how-to-fix-it/?sh=41a71b0b63f0.

Kulick, C., Perry, E. and Bourhis, A. (2000), "Ironic evaluation processes: effects of thought suppression on evaluations of older job applicants", *Journal of Organizational Behaviour*, Vol. 21 No. 6, pp. 689-711.

Kumar, R. and 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*, Vol. 7 No. 3, p. 487.

Lai, C.K., Skinner, A.L., Cooley, E., Murrar, S., Brauer, M., Devos, T. and Nosek, B.A. (2016), "Reducing implicit racial preferences: II. Intervention effectiveness across time", *Journal of Experimental Psychology: General*, Vol. 145 No. 8, pp. 1001-1016.

Legault, L., Gutsell, J.N. and Inzlicht, M. (2011), "Ironic effects of antiprejudice messages how motivational interventions can reduce (but also increase) prejudice", *Psychological Science*, Vol. 22 No. 12, pp. 1472-1477.

Levy, S.R. and Killen, M. (2008), *Intergroup Attitudes and Relations in Childhood through Adulthood*, Oxford University Press, New York, NY.

Lilienfeld, S.O. (2017), "Microaggressions: strong claims, inadequate evidence", *Perspectives on Psychological Science*, Vol. 12 No. 1, pp. 138-169.

Lipman, J. (2018), "How diversity training infuriates men and fails women", *Time Magazine*, Vol. 191 No. 4, pp. 17-19.

Luo, C., Sanger, N., Singhal, N., Pattrick, 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. and 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*, Vol. 24, 100442.

Macrae, C.N., Bodenhausen, G.V., Milne, A.B. and Jetten, J. (1994), "Out of mind but back in sight: stereotypes on the rebound", *Journal of Personality and Social Psychology*, Vol. 67 No. 5, p. 808.

Marteau, T., Sowden, A. and Armstrong, D. (1998), "Implementing research findings into practice: beyond the information deficit model", in Haines, A. and Donald, A. (Eds), *Getting Research Findings into Practice*, BMJ Books, London.

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*, Vol. 63, pp. 153-158.

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", World Wide Web, available at: http//seamonkey.ed.asu.edu/~mei- saac/emc598gunpacking.html (accessed 14 December 2005).

Miller, S. (2001), "Public understanding of science at the crossroads", *Public Understanding of Science*, Vol. 10 No. 1, pp. 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*, Vol. 65, pp. 469-485, doi: 10.1037/0022-3514.65.3.469.

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

Monteith, M.J., Ashburn-Nardo, L., Voils, C.I. and Czopp, A.M. (2002), "Putting the brakes on prejudice: on the development and operation of cues for control", *Journal of Personality and Social Psychology*, Vol. 83, pp. 1029-1050.

Moss-Racusin, C.A., van der Toorn, J., Dovidio, J.F., Brescoll, V.L., Graham, M.J. and Handelsman, J. (2014), "Scientific diversity interventions", *Science*, Vol. 343 No. 6171, pp. 615-616.

Naff, K. and Kellough, J.E. (2007), "Ensuring employment equity: are federal diversity programs making a difference?", *International Journal of Public Administration*, Vol. 26 No. 12, pp. 1307-1336.

Newkirk, P. (2019), *Diversity Inc.: The Failed Promise of a Billion-Dollar Business*, Bold Type Books, New York, NY.

Nickerson, R.S. (1998), "Confirmation bias: a ubiquitous phenomenon in many guises", *Review of General Psychology*, Vol. 2 No. 2, pp. 175-220, doi: 10.1037/1089-2680.2.2.175.

Nordell, J. (2017), "Is this how discrimination ends?", available at: 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, M.I., Sommers, S.R., Apfelbaum, E.P., Pura, N. and Ariely, D. (2006), "Color blindness and interracial interaction: playing the political correctness game", *Psychological Science*, Vol. 17, pp. 949-953.

Odumeru, J.A. and Ogbonna, I.G. (2013), "Transformational vs transactional leadership theories: evidence in literature", *International Review of Management and Business Research*, Vol. 2 No. 2, p. 355.

Paluck, E.L. (2012), "Prejudice and conflict reduction interventions", in Tropp, L. (Ed.), *Oxford Handbook of Intergroup Conflict*, Oxford University Press.

Paluck, E.L. and Green, D.P. (2009), "Prejudice reduction: what works? A review and assessment of research and practice", *Annual Review of Psychology*, Vol. 60, pp. 339-367.

Paluck, E.L., Porat, R., Clark, C.S. and Green, D.P. (2021), "Prejudice reduction: progress and challenges", *Annual Review of Psychology*, Vol. 72, pp. 14.1-14.28.

Pauker, K., Ambady, N. and Apfelbaum, E.P. (2010), "Race salience and essentialist thinking in racial stereotype development: racial stereotype development", *Child Development*, Vol. 81, pp. 1799-1813, doi: 10.1111/j.1467-8624.2010.01511.x.

Pendry, L.F., Driscoll, D.M. and Field, S.C. (2007), "Diversity training: putting theory into practice", *Journal of Occupational and Organizational Psychology*, Vol. 80, pp. 27-50, doi: 10.1348/096317906X118397.

Plant, E.A. and Devine, P.G. (1998), "Internal and external motivation to respond without prejudice", *Journal of Personality and Social Psychology*, Vol. 75, pp. 811-832.

Plant, E.A. and Devine, P.G. (2009), "The active control of prejudice: unpacking the intentions guiding control efforts", *Journal of Personality and Social Psychology*, Vol. 96, pp. 640-652.

Pohl, R.F. (2004), *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory*, Psychology Press.

Reggev, N., Chowdhary, A. and Mitchell, J.P. (2021), "Confirmation of interpersonal expectations is intrinsically rewarding", *Social Cognitive and Affective Neuroscience*, Vol. 16 No. 12, pp. 1276-1287.

Rossen, I., Hurlstone, M.J. and Lawrence, C. (2016), "Going with the grain of cognition: applying insights from psychology to build support for childhood vaccination", *Frontiers in Psychology*, Vol. 7, p. 1483.

Saad, C., Gibbs, W.C., Cox, W.T.L., Devine, P.G., Gehlbach, H., McIntyre, J. and Xie, B. (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. and Omer, S.B. (2013), "A systematic review of interventions for reducing parental vaccine refusal and vaccine hesitancy", *Vaccine*, Vol. 31 No. 40, pp. 4293-4304.

Sales, B.D. and Folkman, S.E. (2000), *Ethics in Research with Human Participants*, American Psychological Association, Washington, DC.

Schleider, J. and 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*, Vol. 59 No. 2, pp. 160-170, doi: 10.1111/jcpp.12811.

Schultz, P.W. (2002), "Knowledge, information, and household recycling: examining the knowledge-deficit model of behavior change", in Dietz, T. and Stern, P. (Eds), *Education, Information, and Voluntary Measures in Environmental Protection*, National Academy of Sciences, Washington, DC, pp. 67-82.

Shelton, J.N. and Richeson, J.A. (2005), "Intergroup contact and pluralistic ignorance", *Journal of Personality and Social Psychology*, Vol. 88 No. 1, p. 91.

Siden, J.Y., Carver, A.R., Meja, O.O. and Townsel, C.D. (2022), "Reducing implicit bias in maternity care: a framework for action", *Women's Health Issues*, Vol. 32 No. 1, pp. 3-8.

Simis, M.J., Madden, H., Cacciatore, M.A. and Yeo, S.K. (2016), "The lure of rationality: why does the deficit model persist in science communication?", *Public Understanding of Science*, Vol. 25 No. 4, pp. 400-414.

Smith, T.C. (2017, July), "Vaccine rejection and hesitancy: a review and call to action", in *Open Forum Infectious Diseases*, Oxford University Press, Vol. 4, No. 3.

Stewart, T.L., La Duke, J.R., Bracht, C., Sweet, B.A.M. and 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*, Vol. 33, pp. 1898-1921.

Stewart, T.L., Latu, I.M., Kawakami, K. and Myers, A.C. (2010), "Consider the situation: reducing automatic stereotyping through situational attribution training", *Journal of Experimental Social Psychology*, Vol. 46 No. 1, pp. 221-225.

Suldovsky, B. (2017), "The information deficit model and climate change communication", *Oxford Research Encyclopedia of Climate Science*, Oxford University Press, Oxford.

Sweeny, K., Melnyk, D., Miller, W. and Shepperd, J.A. (2010), "Information avoidance: who, what, when, and why", *Review of General Psychology*, Vol. 14 No. 4, pp. 340-353.

Tang, A. and Huang, M.Y. (2021), " *The problem with implicit bias trainings*", Business Insider, available at: https://www.businessinsider.com/implicit-bias-trainings-imperfect-should-we-stop-them-2020-7.

Tønnesen, P. (2002), "How to reduce smoking among teenagers", *European Respiratory Journal*, Vol. 19, pp. 1-3.

Trevors, G. and Duffy, M.C. (2020), "Correcting COVID-19 misconceptions requires caution", *Educational Researcher*, Vol. 49 No. 7, pp. 538-542.

Uhlmann, E.L. and 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*, Vol. 104 No. 2, pp. 207-223.

Wenzlaff, R.M., Wegner, D.M. and Roper, D.W. (1988), "Depression and mental control: the resurgence of unwanted negative thoughts", *Journal of Personality and Social Psychology*, Vol. 55, pp. 882-892, doi: 10.1037/0022-3514.55.6.882.

Wilton, L., Apfelbaum, E. and Good, J. (2019), "Valuing differences and reinforcing them: multiculturalism increases race essentialism", *Social Psychological and Personality Science*, Vol. 10 No. 5, pp. 681-689.

Wood, W. (2017), "Habit in personality and social psychology", *Personality and Social Psychology Review*, Vol. 21 No. 4, pp. 389-403.

Wood, W. and Neal, D.T. (2016), "Healthy through habit: interventions for initiating and maintaining health behavior change", *Behavioral Science and Policy*, Vol. 2 No. 1, pp. 71-83.

Wynne, B. (1992), "Misunderstood misunderstanding: social identities and public uptake of science", *Public Understanding of Science*, Vol. 1 No. 3, p. 281.

About the author
William Taylor Laimaka Cox, PhD, is a scientist–practitioner with 20 years of experience in the realm of social justice. His work all serves the ultimate goal of understanding and reducing the injustice, human suffering, and disparities that arise from stereotyping and prejudice. His nonprofit, Inequity Agents of Change, is dedicated to widespread dissemination of evidence-based methods to create lasting, meaningful change. They provide training and resources to individuals and organizations, harnessing the science of cognitive-behavioral change to empower people as agents of change to reduce bias, create inclusion, and promote equity. Learn more at www.biashabit.com William Cox can be contacted at: william.cox@biashabit.com