

Implicit-Bias Remedies: Treating Discriminatory Bias as a Public-Health Problem

May 17, 2022

Psychological Science in the Public Interest (Volume 23, Number 1)

Read the Full Text ([PDF](#), [HTML](#))



Implicit bias refers to thoughts, attitudes, or stereotypes that can be measured indirectly and can operate without awareness. Such measurements have been associated with discriminatory judgments and behaviors, prompting efforts by many researchers to understand how to overcome implicit biases. Some researchers have suggested interventions at the individual level to weaken or eradicate implicit biases. Others have recommended training programs administered to groups to overcome biases more broadly, including implicit ones. Both approaches appear to be inefficient. A third, potentially more-promising approach involves intervening at the institutional level and treating implicit bias as a public-health problem.

In this issue of *Psychological Science in the Public Interest* ([Volume 23, Issue 1](#)), Anthony G. Greenwald, Nilanjana Dasgupta, John F. Dovidio, Jerry Kang, Corinne A. Moss-Racusin, and Bethany A. Teachman elaborate on public-health strategies that have been successful in reducing implicit bias and accompanying discrimination. The researchers also provide recommendations to guide organizations aiming to deal with biases for which these organizations haven't found solutions yet.

What is implicit bias?

“Implicit bias was developed in psychology as a label for mental associations that, when triggered by demographic characteristics such as race, gender, or age, can influence judgment and behavior,” Greenwald and colleagues write. Implicit bias can lead to discriminatory outcomes even when those who perpetrate discrimination are not aware of their biases, much less how those biases may guide their judgment and behavior.



Measuring Implicit Bias: The Implicit Association Test

Implicit biases can be measured indirectly by tools such as the [Implicit Association Test](#) (IAT; Greenwald et al., 1998), which measures the implicit associations individuals make among categories of stimuli simply by having respondents press a key to classify exemplars into their objective categories. For example, to measure implicit gender-science stereotypes, researchers could use stereotype-congruent blocks and stereotype-incongruent blocks. In the first (congruent) scenario, they would ask individuals to press the left key when they see a female name (e.g., “Anna”) on the screen and the right key when they see a male name (e.g., “Ben”), and the left key for words associated with family (e.g., “home”) and the right key for words associated with science (e.g., “lab”). In the second scenario (stereotype-incongruent blocks), they would ask individuals to press the left key to categorize male names and family words, and the right key to categorize female names and science words. The implicit bias is measured by the IAT’s *D* score, computed with a scoring algorithm that considers the latency difference between the two types of blocks. To simplify the example, if individuals more quickly use the same key to categorize family and female than to categorize science and female, they are showing an implicit bias in which they implicitly associate females with family but not with science.

Fully understanding the concept of implicit bias requires correcting common misunderstandings that the public and some scientists appear to have. Greenwald and colleagues compiled those misunderstandings and their corrections:

Misunderstanding 1: The Implicit Association Test (IAT) and other indirect measures assess prejudice and racism. **Correction:** Indirect measures capture associative knowledge about groups, not hostility toward them.

Misunderstanding 2: Implicit measures predict spontaneous (automatic) behavior but do not predict deliberate (controlled, rational) behavior. **Correction:** Implicit measures predict both spontaneous and deliberate behavior.

Misunderstanding 3: Implicit and explicit biases are unrelated to each other. **Correction:** Implicit and explicit biases are almost invariably positively correlated.

Misunderstanding 4: It is scientifically established that long-established implicit biases are durably modifiable. **Correction:** With only occasional exceptions, long-established biases have not been shown to be durably modifiable in experimental attempts.

Misunderstanding 5: Group-administered procedures (often called *antibias* or *diversity training*) that are widely offered to reduce implicit race (and other) biases are effective methods of mitigating discriminatory bias. **Correction:** Scholarly reviews of the effectiveness of group-administered antibias or diversity-training methods have not found convincing evidence for their mental or behavioral debiasing effectiveness.

Greenwald and colleagues also pinpoint three aspects of implicit bias that can be useful for practitioners

working to remediate problems likely to be influenced by implicit bias. Empirical research supports but does not entirely establish these aspects.

1. Causation: Implicit bias is a possible cause of discriminatory behavior.
2. Pervasiveness: Implicit bias is considerably more widespread than is generally expected.
3. Awareness: Implicit bias may produce discriminatory behavior in persons who are unaware of their bias.

Evaluating the remedies for implicit bias

Researchers and practitioners have designed and evaluated different procedures to reduce implicit biases. These include the following:

Experimental interventions, tested in experimental studies. These include single-session experimental interventions (e.g., exposure to counterstereotypical exemplars, appeals to egalitarian values); interventions based on the contact hypothesis, which posits that contact between members of two groups leads to increased liking or decreased disliking between the two groups; and multisession laboratory interventions.

Interventions in field settings, in which the effectiveness of implicit-bias interventions is investigated in real-life settings. Few studies fall into this category.

Large-scale field experiments that include many participants. Some have been implemented in the workplace.

Group-administered trainings administered by large organizations, usually with three components: defining implicit bias as a source of unintended discrimination, describing the pervasiveness of implicit biases, and advocating for remedial strategies.

Interventions targeting attitudes and self-concepts related to clinical disorders. Tested in clinical settings, many interventions target implicit associations with the self (e.g., associations of the self with death [vs. life] to capture suicidality) or with appetitive or aversive stimuli (e.g., associations with alcohol and approach vs. avoidance).

However, Greenwald and colleagues' review of these different types of interventions indicates that both the individual treatment interventions and the group-administered training programs lack established methods that diminish implicit biases in the long term and reduce discriminatory consequences of biases. Although this is a disappointing conclusion, the authors offer a set of strategies that may be more effective. These strategies, based on methods that have been successful in public health, include the following:

Preventive measures, which are designed to disable the path from implicit biases to discriminatory outcomes.

Disparity-finding methods, which aim to discover disparities and how/who to fix them and have the advantage of being useful in remediation for both implicit and systemic biases.

Implicit bias as a public–health problem: Prevention matters

“Mental debiasing and group-administered training are *curative* remedies, aimed at altering mental structures or processes believed to be responsible for discriminatory bias. In contrast, most public-health strategies are directed at persons who are not yet in a condition that calls for curative treatment. Public-health strategies are very often designed to *prevent* rather than to cure,” write Greenwald and colleagues. Thus, treating implicit bias as a public-health problem could lead to interventions that focus on preventing, finding, and measuring inequalities so they can be addressed. In fact, the American Public Health Association (APHA) started reporting, especially after the murder of George Floyd in May 2020, on race discrimination as a public-health problem.

Greenwald and colleagues suggest that public-health remedies for discriminatory bias can be preventive (harm avoiding), governmental (e.g., health laws, regulations, and mandates), and reparative (damage fixing), which differ from the curative interventions reviewed earlier (i.e., mental debiasing and group-administered training). Focusing on preventive remedies, the authors propose **decision blinding**, which involves procedures that prevent a decision-maker from knowing the demographic characteristics of a person or group being evaluated; and **discretion elimination**, which includes procedures in which the bases for deciding are restricted to non-demographic decision-relevant information. Decision blinding and discretion elimination may be achieved by using a set of procedures involving structured interviews, objective testing of aptitudes and skills, objective scoring of written materials, and artificial intelligence to make unbiased decisions.

Finally, to put their conclusions to work, Greenwald and colleagues propose four practical strategies that organizations can implement to reduce discrimination resulting from implicit biases:

1. Make disparity finding a standard practice.
2. Prioritize the use and development of strategies based on bias prevention (vs. strategies based on individual and group debiasing).
3. Use caution regarding remedies described as “training.” Training might improve education about biases, but it is unlikely to change implicit associations that cause discrimination.
4. Integrate diversity, equity, and inclusion efforts into the organizational structure.

Implicit Bias Is a Public-Health Problem, and Hearts and Minds Are Part of the Solution

By Michael A. Olson and Laura J. Gill,
University of Tennessee

Read the Full Text ([PDF](#), [HTML](#))

“Hearts and minds” might contribute to solving public-health problems

In an accompanying commentary, Michael A. Olson and Laura J. Gill (University of Tennessee), argue that interventions to reduce discrimination caused by implicit bias at the individual and group levels (i.e., “hearts and minds” approaches) are as important as preventive strategies at the institutional level. Thus, they encourage the integration of individual-level and public-health perspectives to address implicit bias. Noting some of their disagreements with Greenwald and colleagues, Olson and Gill propose the use of a dual-process framework (motivation and opportunity as determinants [MODE] model). The MODE model considers implicit bias the starting point of the attitude/behavior relation; the activation of bias depends on its strength and lack of awareness of its effects. Motivation and opportunity will then decide whether bias will guide behavior. Hence, the model identifies three key focus areas for interventions: (a) increasing awareness of the impact of bias; (b) increasing motivation to counteract the effects of bias; and (c) increasing opportunity for motivated processes to override implicit ones and thus prevent the effects of bias on behavior.

See related [news release](#).

About the authors ([PDF](#), [HTML](#))

Feedback on this article? Email apsobserver@psychologicalscience.org or comment below.