Curbing implicit bias: what works and what doesn't

Psychologists have yet to find a way to diminish hidden prejudice, but they do have strategies for thwarting discrimination

By Betsy Mason 06.04.2020

Aquarter-century ago, social psychologist Anthony Greenwald of the University of Washington developed a test that exposed an uncomfortable aspect of the human mind: People have deep-seated biases of which they are completely unaware. And these hidden attitudes — known as implicit bias — influence the way we act toward each other, often with unintended discriminatory consequences.

Since then, Greenwald and his main collaborators, Mahzarin Banaji and Brian Nosek, have used the implicit association test to measure how fast and accurately people associate different social groups with qualities like good and bad. They have developed versions of the test to measure things such as unconscious attitudes about race, gender stereotypes and bias against older people. Those tests have revealed just how pervasive implicit bias is. (Project Implicit offers public versions of the tests on its website here.)

Psychologist Anthony Greenwald

University of Washington

The researchers' work has also shown how much implicit bias can shape social behavior and decision-making. Even people with the best intentions are influenced by these hidden attitudes, behaving in ways that can create disparities in hiring practices, student evaluations, law enforcement, criminal proceedings — pretty much anywhere people are making decisions that affect others. Such disparities can result from bias against certain groups, or favoritism toward other ones. Today, implicit bias is widely understood to be a cause of unintended discrimination that leads to racial, ethnic, socioeconomic and other inequalities.

Discussions around the role of racism and implicit bias in the pattern of unequal treatment of racial minorities by law enforcement are intensifying following a roster of high-profile cases, most recently the killing of George Floyd. Floyd, an unarmed black man, died in Minneapolis last month after a white police officer pressed his knee into Floyd's neck for nearly nine minutes.

As awareness of implicit bias and its effects has increased, so has interest in mitigating it. But that is much harder to do than scientists expected, as Greenwald told an audience in Seattle in February at the annual meeting of the American Association for the Advancement of Science. Greenwald, coauthor of an overview on implicit bias research

in the 2020 *Annual Review of Psychology*, spoke with *Knowable Magazine* about what does and doesn't work to counter the disparities that implicit bias can produce.

This conversation has been edited for length and clarity.

How do you test for associations that people aren't aware they have?

The first implicit association test I created was one involving the names of flowers and insects, and words meaning things pleasant or unpleasant. You had to use left and right hands to classify them, tapping on a keyboard as they appeared on the screen. It was a very easy task when you had to use the right hand for both pleasant words and flower names, and the left hand for unpleasant words and insect names, because we typically think of flowers as pleasant and insects as unpleasant.

But then the task is switched to force the opposite associations — one hand for insect names and pleasant words, and the other hand for flower names and unpleasant words. When I first tried that reversed form, my response time was about a third of a second slower compared to the first version. And in psychological work where you're asking people to respond rapidly, a third of a second is like an eternity, indicating that some mental processes are going on in this version of the test that are not going on in the other.

Then I replaced the flowers and insects with first names of men and women that are easily classified as European American or African American. For me, giving the same response to pleasant words and African American names took an eternity. But when it was the European American names and pleasant words with one hand, and the African American names and the unpleasant words with the other hand, that was something I could zip through. And that was a surprise to me. I would have described myself at that point as someone who is lacking in biases or prejudices of a racial nature. I probably had some biases that I would confess to, but I actually didn't think I had that one.

How widespread is implicit bias?

That particular implicit bias, the one involving black-white race, shows up in about 70 percent to 75 percent of all Americans who try the test. It shows up more strongly in white Americans and Asian Americans than in mixed-race or African Americans. African Americans, you'd think, might show just the reverse effect — that it would be easy for them to put African American together with pleasant and white American together with unpleasant. But no, African Americans show, on average, neither direction of bias on that task.

Most people have multiple implicit biases they aren't aware of. It is much more widespread than is generally assumed.

Is implicit bias a factor in the pattern of police violence such as that seen in the killing of George Floyd on May 25, which sparked the ongoing protests across the country?

The problems surfacing in the wake of George Floyd's death include all forms of bias, ranging from implicit bias to structural bias built into the operation of police departments, courts and governments, to explicit, intended bias, to hate crime.

The best theory of how implicit bias works is that it shapes conscious thought, which in turn guides judgments and decisions. The ABC News correspondent Pierre Thomas expressed this very well recently by saying, "Black people feel like they are being treated as suspects first and citizens second." When a black person does something that is open to alternative interpretations, like reaching into a pocket or a car's glove compartment, many people — not just police officers — may think first that it's possibly dangerous. But that wouldn't happen in viewing a white person do exactly the same action. The implications of conscious judgment being shaped in this way by an automatic, implicit process of which the perceiver is unaware can assume great importance in outcomes of interactions with police.

Do the diversity or implicit bias training programs used by companies and institutions like Starbucks and the Oakland Police Department help reduce bias?

I'm at the moment very skeptical about most of what's offered under the label of implicit bias training, because the methods being used have not been tested scientifically to indicate that they are effective. And they're using it without trying to assess whether the training they do is achieving the desired results.

I see most implicit bias training as window dressing that looks good both internally to an organization and externally, as if you're concerned and trying to do something. But it can be deployed without actually achieving anything, which makes it in fact counterproductive. After 10 years of doing this stuff and nobody reporting data, I think the logical conclusion is that if it was working, we would have heard about it.

Can you tell us about some of the approaches meant to reduce bias that haven't worked?

I'll give you several examples of techniques that have been tried with the assumption that they would achieve what's sometimes called debiasing or reducing implicit biases. One is exposure to counter-stereotypic examples, like seeing examples of admirable scientists or entertainers or others who are African American alongside examples of whites who are mass murderers. And that produces an immediate effect. You can show that it will actually affect a test result if you measure it within about a half-hour. But it

was recently found that when people started to do these tests with longer delays, a day or more, [any beneficial effect appears to be gone].

Other strategies that haven't been very effective include just encouraging people to have a strong intention not to allow themselves to be biased. Or trainers will suggest people do something that they may call "thinking slow" or pausing before making decisions. Another method that has been tried is meditation. And another strategy is making people aware that they have implicit biases or that implicit biases are pervasive in the population. All these may seem reasonable, but there's no empirical demonstration that they work.

It's surprising to me that making people aware of their bias doesn't do anything to mitigate it. Why do you think that is?

I think you're right, it is surprising. The mechanisms by which our brains form associations and acquire them from the cultural environment evolved over long periods of time, during which people lived in an environment that was consistent. They were not actually likely to acquire something that they would later have to unlearn, because the environment wasn't going to change. So there may have been no evolutionary pressure for the human brain to develop a method of unlearning the associations.

I don't know why we have not succeeded in developing effective techniques to reduce implicit biases as they are measured by the implicit association test. I'm not prepared to say that we're never going to be able to do it, but I will say that people have been looking for a long time, ever since the test was introduced, which is over 20 years now, and this hasn't been solved yet.

Is there anything that does work?

I think that a lot can be achieved just by collecting data to document disparities that are occurring as a result of bias. And maybe an easy example is police operations, although it can be applied in many settings. Most police departments keep data on what we know as profiling, though they don't like to call it that. It's what happens in a traffic stop or a pedestrian stop — for example, the stop-and-frisk policy that former New York City Mayor Michael Bloomberg has taken heat for. The data of the New York City Police Department for stops of black and white pedestrians and drivers were analyzed, and it was very clear that there were disparities.

Once you know where the problem is that has to be solved, it's up to the administrators to figure out ways to understand why and how this is happening. Is it happening in just some parts of the city? Is it that the police are just operating more in Harlem than in the white neighborhoods?

And once you know what's happening, the next step is what I call discretion elimination. This can be applied when people are making decisions that involve subjective judgment about a person. This could be police officers, employers making hiring or promotion decisions, doctors deciding on a patient's treatment, or teachers making decisions about students' performance. When those decisions are made with discretion, they are likely to result in unintended disparities. But when those decisions are made based on predetermined, objective criteria that are rigorously applied, they are much less likely to produce disparities.

Is there evidence that discretion elimination works?

What we know comes from the rare occasions in which the effects of discretion elimination have been recorded and reported. The classic example of this is when major symphony orchestras in the United States started using blind auditions in the 1970s. This was originally done because musicians thought that the auditions were biased in favor of graduates of certain schools like the Juilliard School. They weren't concerned about gender discrimination.

But as soon as auditions started to be made behind screens so the performer could not be seen, the share of women hired as instrumentalists in major symphony orchestras rose from around 10 percent or 20 percent before 1970 to about 40 percent. This has had a major impact on the rate at which women have become instrumentalists in major symphony orchestras.

But these data-collection and discretion-elimination strategies aren't commonly used?

Not nearly as often as they could. For example, instructors can usually arrange to grade almost anything that a student does without knowing the identity of the student. In an electronic age when you don't learn to recognize people's handwriting, instructors can grade essays without the students' names on them. I used that approach when I was last grading undergraduates in courses. It's easy to use, but it's often not used at all.

And in many other circumstances it is possible to evaluate performance without knowing the identity of the person being evaluated. But employers and others rarely forgo the opportunity to know the identity of the person they're evaluating.

Can artificial intelligence play a role?

People are starting to apply artificial intelligence to the task by mining historical records of past employment decisions. This is a way of taking the decisions that involve human discretion and putting them into the hands of a machine. The idea is to develop

algorithms that identify promising applicants by matching their qualities to those of past applicants who turned out to be successful employees.

I think it's a great thing to try. But so far, efforts with AI have not succeeded, because the historical databases used to develop the algorithms to make these decisions turn out to be biased, too. They incorporate the biases of past decision-makers. One example is how biases affect [facial-recognition technology](#), which inadvertently categorizes African American faces or Asian faces as criminal more often than white faces.

This is a problem that computer scientists are trying to cope with, but some of the people in AI that I have spoken to seem not so optimistic that this will be at all easy to do. But I do think that ultimately — and it might take a while — the biases may be expunged more readily from AI decision algorithms than from human decision-making.

Could more be done at the level of an individual company or department?

To help prevent unintended discrimination, the leaders of organizations need to decide to track data to see where disparities are occurring. When they discover disparities, they need to try to make changes and then look at the next cycle of data to see if those changes are improving things.

Obviously, it's easier for them not to do those things. In some cases there's a cost to doing them. And they may think it's like opening up Pandora's box if they look closely at the data. I think this is true of many police departments. They're bound to find things that they'd rather not see.