The summer of 2020 witnessed perhaps the largest protests in American history in response to police and vigilante brutality against the black community. New protests are still erupting every time another suppressed video, such as of Daniel Prude, surfaces, or another killing, such as Breonna Taylor’s, goes unpunished. As communities demand meaningful reform, the point—or pointlessness—of “implicit bias training” takes on renewed urgency. Implicit bias trainings aim to raise awareness about the unwitting or unwilling prejudices and stereotypes that shape our habits of thinking, feeling, and navigating through the social world. These trainings have been widely adopted by businesses, schools, and law enforcement agencies. Do they make any difference?

Although I conduct implicit bias trainings myself (including for courts, judges, police, and attorneys), I share many critics’ concerns. Many trainings are too brief and oversimple, and too often their real function is to permit organisations to “check a box” to protect against litigation, rather than to spark real change. But “implicit bias training” is just another way of saying “education about implicit bias,” and, like all kinds of education, it can be done well or poorly. If implicit bias is one important piece of a large and complex puzzle, then education about it—when done right—should have a meaningful role to play in helping us understand ongoing inequities and enact reforms.

First, however, we might step back and consider the function of modern-day education more broadly. The worry that implicit bias training just serves a box-checking function is a specific instance of a farther-reaching and longer-standing concern. The concern is that formal education is less about acquiring useful information and skills than it is about signalling. For example, given that we forget the vast majority of what we learn in school, what purpose does a high GPA really serve other than to signal to potential employers that we are hard-working, conscientious, even sycophantic “team players” who eagerly obey our superiors’ instructions? In other words, the message your Bachelor’s degree sends to potential bosses is: you’ll do what they tell you.

Many trainings are too brief and oversimple

In The Case Against Education, Bryan Caplan uses a simple thought experiment to illustrate the point: “you can have either a Princeton education without a diploma, or a Princeton diploma without an education. Which gets you further...?” (The question is reminiscent of Glaucon and Adeimantus’ challenges to Socrates in the Republic: is it better to be perfectly good person with a
Thoughts

perfectly evil reputation, or a perfectly evil person with a perfectly good reputation?) So which would you prefer in Caplan’s example: Ivy League education minus reputation, or Ivy League reputation minus education? If this is something you have to mull over even for a moment, rather than immediately opting for the former, the implications for academia are damning. The signalling value of the degree would rival all the skills and knowledge graduates (are supposed to) have learned. (Anecdotally, my students seem split between the two options.)

With these general questions about educational signalling in view, which signals might be sent by implicit bias training in particular? For one thing, institutions can use these sessions to convey to their employees, stakeholders, consumers, and would-be litigators that they are doing their part to fight bias. For another, participants who attend these sessions can signal to their superiors and each other that they care about racism, sexism, and so on. In these ways, implicit bias training isn’t so much about signalling the dissemination or acquisition of knowledge. It’s about signalling virtue. Once the box is checked, however, if there are no further mechanisms for holding individuals and institutions accountable, everyone can go back to business as usual.

I fear the foregoing describes a great deal of what goes on under the guise of implicit bias training, but the first point to make...
is that these observations about signalling have little if anything to do with the content or nature of the training itself. Implicit bias training could be highly effective or completely ineffective (to the point that people could “attend” these trainings without paying a moment’s attention or absorbing a single lesson), and yet it could still serve the same box-checking functions.

The immediate question to ask, then, is how many of these justified concerns about implicit bias training should be directed at it rather than at the systems and structural constraints within which it is embedded. How many of the apparent shortcomings of implicit bias training reflect problems of supply, regarding the knowledge being produced and shared by social scientists and diversity trainers, and how many shortcomings instead reflect problems of demand, regarding the institutions within which these trainings occur and the attitudes and expectations of those who implement and participate in them? Suppose the members of a Faculty Search Committee are required to attend bias training before crafting their job ad and embarking on their search, and the Search Committee Chair tells his colleagues, “I know this training is a pain, but we just have to do it to get HR off our back.” Then suppose no further procedures hold the committee responsible for following through on an equitable and unbiased search. If the committee fails to make any substantive changes to its search practices, how much is this a failure of the implicit bias training per se? The deeper problem would be the broader set of norms, attitudes, and procedures wrapped around the training.

It’s plausible, then, that some of these justified concerns float free of what actually happens during bias training. Yet some do worry that the content of implicit bias training is inherently problematic. Given the resurgence in overt bigotry and intergroup hostility, why are we talking about implicit bias at all? Explicit bias seems the more serious problem. But it is a mistake to portray explicit and implicit forms of bias as opposing explanations of social injustice. In fact they feed into each other. Research finds, for example, that implicit biases become explicit when leaders promote discriminatory norms and values.

A related worry is that the concept of implicit bias is irredeemably individualistic, locating the sources of injustice in our hearts and minds rather than in wider social structures. It is true that popular discussions of implicit bias are simplistic and individualistic, but popular discussions of almost everything are simplistic and individualistic. Our collective attraction (especially in Western, Educated, Industrialised, Rich, and Democratic (WEIRD) parts of the world) to individualistic narratives is well-documented, as is our cognitive temptation to explain events in terms of actors’ internal traits and motivations rather than in terms of their situations. By contrast, it is a standard and widely practiced feature of these trainings to stress that the roots of bias reside in factors outside individual minds, such as in mass me-
about implicit bias is not by itself going to make people less biased. Nor is educating people about the world’s entrenched structural injustices by itself going to make the world more just! Changing the world for the better requires action. We do well here to remember Aristotle’s criticisms of those who “take refuge in theory” rather than invest in building up ethical habits and embodying just norms:

It is well said, then, that it is by doing just acts that the just man is produced, and by doing temperate acts the temperate man; without doing these no one would have even a prospect of becoming good. But most people do not do these, but take refuge in theory and think they are being philosophers and will become good in this way, behaving somewhat like patients who listen attentively to their doctors, but do none of the things they are ordered to do. As the latter will not be made well in body by such a course of treatment, the former will not be made well in soul by such a course of philosophy.

Bias education needs to become resistance training

With Aristotle’s admonishment of the politically inert in mind, implicit bias training should be seen as a failure in the event that it demotivates us, making us feel less rather than more likely to pursue these questions further. It is a failure if participants become frozen with fear that their
Thoughts

efforts to enact change will blow up in their faces. And it is certainly a failure if individuals leave smugly thinking, “Now that I’ve shown up to this, I’ve done my part, and I can relax.” In all of these ways, implicit bias training can do more harm than good.

Conversely, implicit bias training is a success in the event that participants come away with a little extra motivation to tackle injustice, and with concrete ideas for what to do next. Where can they go to learn more? How can they link up with likeminded folks already working to bring about change? What sorts of experiments-in-living or experiments-in-the-lab might they try out to combat bias?

We must, moreover, understand the notion of “combating bias” broadly. It doesn’t just mean becoming less biased in our concrete interactions with other individuals. Combating bias also means building up what I call structure-facing habits and virtues. Effective implicit bias training ought to cultivate a multidirectional orientation, guiding our attention both to the biased ways we interpret and react to other individuals, and
to the unjust structures that surround us. And it ought to motivate us to take action in response. In short, bias education needs to become resistance training. Becoming less biased means becoming the sort of person who reliably recognises, and resists, the unjust status quo.

How might we revise implicit bias training to better serve this aim? Apart from rethinking the material we teach in these sessions, we do well to consider the way the training is structured. We might take a page from Pedagogy of the Oppressed, where Paulo Freire explains how basic literacy education should not just be about “depositing” information about letters, phonemes, and words in learner’s heads. It should also, he argues, turn learners away from passive acquiescence to unjust social realities and toward being more critically and socially engaged. Arguably, there is nothing inherent to teaching the alphabet that cultivates political resistance rather than adherence to the status quo. Freire argues that how literacy is taught makes the difference. Along with numerous recommendations for building a curriculum that resonates with learners’ lived experiences, he emphasises that education should be dialogical rather than lecture-based, and above all that it should be anti-hierarchical. That is to say that rather than training students to sit and submissively absorb what they’re told from the all-knowing instructor, they should be actively engaged. Creating this egalitarian and critical pedagogical environment requires that the teachers come to learn and the students come to teach.

Now, I would not claim that my approach to implicit bias training regularly embodies the ideals of radical pedagogy. I’ve given my fair share of one-man-show lectures. With only 60 minutes, and dozens or hundreds of people in the room (or in the Zoom, these days), the context does not afford extensive, reciprocal-learning dialogue. Within such constraints, I don’t think these sessions are worthless, but my hopes for what to achieve are modest at best.

If, however, the widely recognised shortcomings of implicit bias training come down to such artificial constraints, perhaps the problem is not that we’re doing too much implicit bias training, but too little. More extensive, intimate, and interdisciplinary engagements with the subject are called for. Maybe instead of one-page handouts, we need entire textbooks. Maybe instead of one-off workshops, we need complete courses. That, in any event, was the theory behind An Introduction to Implicit Bias: Knowledge, Justice, and the Social Mind, which I co-edited with Erin Beeghly. And it was the rationale behind my new course at Cal Poly Pomona, The Philosophy and Science of Implicit Bias, where the final exam is not to write another paper but to leave the classroom and try to change the world.