Inherent in the idea of democratic society is the possibility of changing one another’s beliefs, attitudes, and behavior, albeit in limited and modest ways. It is difficult to imagine maintaining faith in democracy as a political system without believing that we can sometimes move one another without resorting to force. And yet, in practice, the challenge of changing others’ minds is maddeningly familiar. This is particularly so when it comes to diminishing biases and prejudices.

I begin here with the observation that some people are better at this than others. And so I ask: what kind of skill does it take to be good at changing people’s biases and prejudices? What are the strategies, abilities, or traits associated with being a successful “de-biasing agent?”

Before trying to answer this question, I offer a few caveats and clarifications.

(1) I focus here on the skills individuals might seek, practice, and master. I have in mind situations of interpersonal interaction, perhaps between family members, friends, neighbors, and citizens. In focusing in this way, I do not mean to implicitly criticize other deeply worthwhile anti-discrimination and anti-prejudice activities, such as organizing protests, crafting legislation, establishing case law, amending corporate human resources policies, and militating for changes in public policy. I will return briefly in the conclusion to discuss how the ideas I present here relate to the ongoing debate between proponents of “individual” versus “structural” approaches to social change.

(2) Nor do I mean to imply that the self-regulation of prejudice and stereotyping is unimportant. I have biases and prejudices. It is my responsibility to try to diminish them. Indeed, as I discuss below, reducing prejudice is more a matter of modeling desirable behavior than it is a matter of persuading people to believe what you want them to. If this is right, then skilled de-biasing depends in part upon one’s ability to regulate one’s own prejudices. I have addressed the self-regulation of prejudice elsewhere; here I address what people should consider doing if they want to combat injustice by trying to change the biases their fellow citizens hold. Concomitant to this, it is important to remember that many people want to be fair minded and to act in relatively egalitarian ways. In this sense, it is worth considering how to help each other challenge our biases.
“Bias” is sometimes used in a pejorative sense, sometimes not. In the non-pejorative sense, a bias is a disposition to perceive or cognize things in a particular way, given some trait or feature of one’s identity. Classical musicians might be biased in favor of the classical music radio station in this sense; orthopedic surgeons might be biased against letting their kid play on the jungle gym. In the pejorative sense, biases are inherently unfair, or they obscure the truth in some way. Racists are biased in the pejorative sense. What makes racist beliefs biased is not just that they are skewed by the racist’s own idiosyncratic view of things. Racist beliefs are biased because they are unfair, untrue, or unjustified. They are flawed, whether factually or morally (or both). This pejorative sense of “bias” is the one I use in this chapter. Below I will note some important differences in the psychology underlying the relevant kinds of factual and moral errors.

As I understand them, prejudices are negative attitudes (i.e., evaluative feelings and thoughts, otherwise known as likings and dislikings) about people in virtue of their social group membership; stereotypes are generalizations about the traits of groups; and discrimination is unfair treatment on the basis of social group membership, often driven by prejudice and stereotyping. Despite their distinct meanings, prejudice, stereotyping, and discrimination are fellow-travelers. Elsewhere, I have defended a view that there is a blurry boundary, at best, between these concepts. This issue won’t matter much here, although I will revisit it briefly in Section 39.1.

Regarding “skill,” I intend to stay neutral with respect to theories contending that skill is a form of knowledge-how (Ryle 1949); a particular way of responding to “affordances” (Dreyfus 2002a, 2002b); or a disposition to know certain propositions (Stanley & Williamson 2017). That is, I am here less interested in the nature of skill than I am in the specific skills a successful de-biaser might display. What abilities must such a person possess? What knowledge must she have? What strategies should she use? I operate with the assumption that de-biasing is a skill, however. Like dancing, debating, and drawing well, de-biasing is something one can learn and improve over time with practice. Acquiring the skill involves a combination of adopting the right goals, learning the right information, and simply practicing the relevant actions repeatedly. Doing so enables one to become fluent, that is, able to respond to different kinds of situations in appropriate ways on the fly. Most people, I suspect, can call to mind exemplars of de-biasing, people whom they believe were unusually skilled at it. What is it that these exemplars do?

39.1 Resisting argumentation

First, skilled de-biasers tend to resist the temptation to try to argue people out of their opinions. One reason for this is to avoid being condescending or perceived as holier-than-thou. People tend not to like being told that they are prejudiced.

“Opinions” is a purposely vague term. As I’ll discuss below, argumentation can be successful for changing people’s factual beliefs, but usually much less so for changing people’s attitudes, intentions, and behavior. It suffices to be wary of it as a route for combating prejudice, stereotyping, and discrimination. Of course, not everyone feels the temptation to argue people out of their opinions equally. But feeling it is common, perhaps particularly among intellectuals whose professional currency is ostensibly argumentation.

The fact is, however, that arguments are often ineffective tools. Long-standing laboratory research on belief perseverance, confirmation bias, and motivated reasoning (e.g., Ross et al. 1975; Lord et al. 1979; Kunda 1990; Nickerson 1998; Kahan et al. 2017; Mandelbaum 2019)
Michael Brownstein

shows, at least minimally, how steep the hill to climb is. These streams of research have, unsurprisingly, been applied in contexts involving intergroup attitudes (e.g., Kunda & Sinclair 1999; Ulhmann & Cohen 2005). Activists and canvassers engaged in shifting individuals’ opinions about specific public policies have absorbed these lessons and tend now to shy away from trying to win arguments. And good reasons are given for the ineffectiveness of argumentation in research on the evolutionary psychology of reasoning, the most compelling accounts of which (e.g., Mercier & Sperber 2011, 2017) suggest that reasoning in human beings evolved to facilitate cooperation, group cohesion, and the sustainment of social identity rather than to evaluate the strength of arguments, make good causal inferences, and so on.

Of course, there are important caveats here. Consider research on political beliefs and the so-called “backfire effect.” A series of highly-cited papers suggests that correcting people’s misperceptions about, for example, weapons of mass destruction in Iraq (Nyhan & Reifler 2010), belief in Obamacare “death panels” (Nyhan et al. 2013), and potential health impacts of climate change (Hart & Nisbet 2012) ironically leads people to double-down on their false beliefs about these topics. Telling people their beliefs are false leads them to hold on to their mistaken beliefs more strongly, in other words. Or so the story went. But it turns out that things are more complex than this. More careful and high-powered studies have failed to find evidence that correcting false beliefs leads people to believe the opposite of the truth (Cameron et al. 2013; Wood & Porter 2016). These follow-up studies have led some to suggest that the backfire effect ought to be thrown in the dustbin of history of much-hyped but ultimately illusory psychological phenomena.

The warranted caveat here is that arguments and facts can correct false beliefs, and so in the domain of prejudice and stereotyping, there may be a time and a place for the skilled de-biaser to inform her racist uncle that violent crime is not on the rise in the United States or that China and India (not Mexico) are the sources of the largest number of people currently migrating to the United States.

But more than this is unwarranted, and for two central reasons. First, while the evidence for backfire effects on factual beliefs is weak, there is mixed evidence at best in the above studies that participants have meaningful uptake of true beliefs. In other words, while participants don’t double-down on their false beliefs in the way the original studies suggested, moving in the opposite direction of truth as a result of being given the facts, participants also don’t move consistently in the direction of the truth once they have the facts. The dynamics of this process—the “continued-influence effect” (Lewandowsky et al. 2012)—are not entirely understood. Repetition of factual corrections, presenting information in worldview-affirming ways, and other tactics appear to diminish it (Lewandowsky et al. 2012; see also Anderson 1982). Skilled de-biasers will need to learn these, as well as to adumbrate the strategy of “going factual” to contextual and personality variables.

Second, and perhaps more importantly, researchers do find evidence of backfire effects when it comes to intentions and attitudes, as compared to factual beliefs. For example, Nyhan and Reifler (2015) find that correcting the false belief that the flu vaccine gives you the flu reduces people’s belief in this myth but also reduces people’s intent to get vaccinated. And Nyhan and colleagues (2017) find that correcting false claims made by Donald Trump reduces belief in those claims among Trump supporters, but it does not affect their attitudes toward Trump. This is crucial when it comes to combating prejudice, stereotyping, and discrimination. While these are separate constructs, in practice they are nearly inseparable. As a result, changing “opinions” about social groups will nearly inevitably involve changing the kinds of attitudes that do appear susceptible to backfire effects. As Gordon Allport presciently said, “information seldom sticks unless mixed with attitudinal glue” (Allport 1954: 485).
De-biasing, skill, and intergroup virtue

The idea that the backfire effect is complex—that it works in some contexts but not others, for people with certain traits but not others—applies to persuasion in general (and, really, to just about any psychological phenomenon). The experimental literature on persuasion is vast and successes are, of course, well-documented. But while persuasion sometimes works, it is often unclear why it works. Proposed explanations involve the credibility, stature, and even attractiveness of the message-giver, the emotional state and meta-cognitive awareness of the message-receiver, the sheer number and length of arguments in a message and how many times the message is repeated, the concomitant presentation of particular images (e.g., brain scans), and so on (for review see Petty & Briñol). It is crucial to understand why persuasion works, when it does, because (a) in many cases these ostensible moderators cut both ways (e.g., in some studies happy people appear more receptive to persuasive arguments and in other studies they appear more resistant to persuasive arguments) and (b) it is possible that what makes persuasion sometimes effective has nothing to do with the strength of one’s arguments. This is clear when attractive interlocutors are more persuasive than others.

All of this said, it is undeniable that providing people with knowledge about prejudice, stereotyping, and discrimination is crucial for creating social change. Effective attitude change techniques do just this. For example, Patricia Devine’s “habit-breaking” approach to prejudice-reduction focuses on providing people with knowledge about the mechanisms of bias, along with motivation to avoid its effects and strategies for effectively avoiding it. Devine and colleagues’ (2012, 2017) research is noteworthy for creating durable change in participants’ attitudes and behavior, for example, leading to increased hiring of women in STEMM fields. But providing a person with knowledge is not tantamount to persuading them to change their mind. Devine’s approach packages knowledge with motivation and tools in an effort to frame bias as a bad habit, not a false ideology. Other approaches provide knowledge by telling someone, “here is how prejudice has affected me,” without prescribing what message-receivers are supposed to conclude from this. (See Section 39.4 for more on this tactic.)

As I granted above, some contexts may be more conducive than others to changing beliefs, attitudes, and behavior via argumentation. Debate societies and peer-reviewed journals, for example, are ostensibly appropriate venues for persuasion. While it’s unclear how often people change their minds in these contexts, and there are worries about who has access to these venues, persuasion is certainly at home here. There are also other, less formal situations in which persuasion and argument may be appropriate. Some friendships and marriages are built on it! But success in these contexts is likely to be built upon mutual trust and a shared set of experiences, which are precisely what is often lacking in intergroup exchanges. These background conditions are more the exception than the norm. One thing that helps to promote them is what psychologists call “intergroup contact.”

39.2 Intergroup contact

Here’s a truism: doing things cooperatively with other people tends to lead people to like each other. But that this is a truism doesn’t diminish its significance. Promoting cooperative intergroup interaction is at the core of the “contact hypothesis,” which is perhaps the most influential and longest-standing research program on prejudice-reduction in existence. Allport’s (1954) central proposition was that intergroup contact promotes intergroup liking. (Put in less sterile terms: doing things with people unlike yourself leads you to like them.) High-powered, longitudinal studies have demonstrated the salutary effects of intergroup contact. Colette Van Laar and colleagues (2005), for instance, studied the effects of both randomly assigned and voluntary intergroup roommate assignment at UCLA over four years and found that for nearly...
all groups they studied (White, Asian American, Latino, and African American), living with a
member of another group decreased prejudice and, perhaps even more significantly, this effect
generalized to feelings about members of other outgroups. Similar positive effects as a result
of intergroup roommate assignment have been found on implicit racial attitudes (Shook &
Fazio 2008).  

So, at first glance, an obvious answer to the question of how to become a skilled de-biaser
might be to embrace the contact hypothesis. But I will not dwell extensively on this approach.
My reason is that it does not fit naturally within the scope of the question I am asking.
Intergroup contact is essential for changing one’s own prejudices. Moreover, it is an important
goal for legislation and litigation (e.g., Brown v. Board of Education of Topeka). But it is unclear
how an individual could “deploy” intergroup contact as a tactic in the relevant sense. One could
try to persuade people to enter situations in which they will interact with outgroup members.
But in many cases this puts the cart before the horse, because prejudice is often what prevents
people from wanting to engage with others outside their own social group. Intergroup contact
can be conceived of as an end, in this sense, that the skilled de-biaser helps people to embrace,
rather than as a means to that end.

39.3 Personal connections

Here’s another seeming truism: changing people’s minds about things they care about requires
making a personal connection with them. For example, Jeremy Bird, the national field dir-
dector for President Obama’s 2012 re-election campaign, reported, “we trained our volunteers to
connect with voters at the door on a personal and values level, not to talk at them with scripted
talking points.” A recent highly publicized study on reducing transphobia substantiates this
idea. David Broockman and Joshua Kalla (2016) showed that short conversations (approximately
ten minutes) with voters in Miami, Florida, in which canvassers asked voters to recall a time in
which they had been judged negatively for being different and then to relate that feeling to the
experiences of transgender people, significantly increased positive feelings toward transgender
people. The effect was large; it was greater than the average increase in positive feelings toward
gay men and lesbians among Americans between 1998 and 2012. The effect was also common
to both registered Republicans and Democrats, as well as to voters whose feelings toward trans-
genre people started off above and below the average. Most strikingly, the effect of these short
conversations did not diminish after three months. 

Broockman and Kalla’s canvassers didn’t try to convince voters of anything. They simply
identified themselves as working with an LGBT organization; informed voters that they
might have to vote whether to repeal a law protecting transgender people; asked voters
if they had an opinion about the law and to explain their opinion if so; showed them
a short video that defined the term “transgender,” described simple arguments for and
against the law, and identified that they were transgender, if they were; asked the voters to
recall an experience as described above and to relate that experience to the experiences of
transgender people; and then asked the voters whether the conversation had in any way
changed their minds.

It isn’t certain what was responsible for Broockman and Kalla’s canvassers’ success. The
researchers themselves understood the intervention as a way of encouraging effortful active
processing and analogic perspective-taking (i.e., encouraging voters to actively reason about
what it would be like to experience the world as a transgender person). This builds upon pre-
vious successes in lab-based research using these techniques (e.g., Galinsky & Moskowitz 2000).
But as the authors note, the
focus on external validity means we cannot be certain that perspective-taking is responsible for any effects or that active processing is responsible for their duration; being primarily concerned with external validity and seeking to limit suspicion, we did not probe intervening processes or restrict the scope of the conversations as a laboratory study would.

It certainly seems clear that canvassers made some kind of personal connection to voters, compared with controls. Indeed, voters in the control group were simply informed that they may have to vote on a law requiring supermarkets to charge for plastic bags, asked what they thought about this, and then thanked. So, in addition to differences in the content of the discussion across conditions, not only did voters in the control condition not experience the active processing and analogical perspective-taking intervention, but they did not experience any kind of meaningful connection with the canvasser. This leaves open a number of possibilities. Broockman and Kalla’s canvassers may have had success in the experimental condition because they encouraged active processing, because they managed to get voters to take the perspective of transgender people, or because they simply made a personal connection with them, notwithstanding the active processing and perspective-taking (or only because they did all three of these things together). Hopefully future research will replicate Broockman and Kalla’s results as well as consider these possibilities as isolated independent variables. Until then, de-biasers would do well to try all three, much in the mold of Broockman and Kalla’s canvassers.

39.4 Social referents and the perception of norms

I have a speculative suggestion about why Broockman and Kalla’s canvassers were successful. They were effective de-biasers, I suggest, because they changed people’s perceptions of norms. They did this, moreover, because they became what Betsey Levy Paluck calls a “social referent” (Paluck & Shepherd 2012; Tankard & Paluck 2016). Social referents are people with particular influence over other people’s perceptions of norms. If correct, the upshot of this idea is that instead of trying to change people’s beliefs or attitudes directly, successful de-biasers change people’s perceptions of what other people believe and feel.

The idea that de-biasing agents become social referents draws upon a large and growing body of research on the psychology of norms. The basic claim of this body of research is that people’s perceptions of norms are crucial determinants of their social attitudes and behavior. There is robust debate about how exactly to understand this claim, of course (e.g., Bicchieri 2005, 2016; Brennan et al. 2013; Henrich 2015; Kelly & Davis 2018). Much of this debate is orthogonal to the discussion here, although below I will note a few commitments I make to contested positions. Much of the debate homes in on a particular sense of the terms norms and normativity, a sense that is different from the way in which these terms are typically deployed in other disciplines. In the most general sense, norms refer to what is required, allowed, or forbidden (Kelly & Davis 2018). When philosophers refer to norms and normativity, they usually consider the nature of claims about what is required, allowed, or forbidden. Are those claims coherent, justified, universal, etc.? Moral philosophers, for example, try to ground normativity in this sense on various kinds of theoretical foundations (e.g., the promotion of the greatest good for the greatest number, according to utilitarians). In contrast, when economists, sociologists, and anthropologists talk about norms and normativity, they usually refer to what’s normal or prevalent in a polity. Mixed-gender schooling is the norm in some cultures but not in others,
for example. Norms in this sense are just regularities regarding what’s required, allowed, or forbidden; economists and other social scientists aren’t usually interested in whether these regularities are well-justified, for example. Psychologists (in particular, evolutionary psychologists) usually understand norms and normativity in a third sense, one that sits somewhat squarely between the philosopher’s and the economist’s senses. Psychologists tend to study people’s perceptions of what’s required, allowed, or forbidden; how and why these perceptions are formed, in terms of the architecture and evolution of the mind; and how perceptions of norms affect people’s attitudes, beliefs, and behavior in distinctive ways (i.e., different from other influences on their attitudes, beliefs, and behavior). Of course, in reality, there is a lot of disciplinary overlap.

I make these generalizations in order to clarify that it is in the third sense—the psychologist’s sense—that I suggest that successful de-biasing agents become social referents for people’s perceptions of norms.

There are many reasons to think that prejudice, stereotyping, and discrimination are illuminated by norm psychology. An overarching reason is that norm psychology helps to explain how social groups maintain cohesion and, by definition, group cohesion requires a distinction between ingroup members and outgroup members. This is, in some sense, the very distinction underlying prejudice, stereotyping, and intergroup discrimination. But there are more specific reasons too. I’ll identify three.

First, people’s attitudes and behavior are influenced by their perception of what other people think it’s appropriate or inappropriate to do, and this effect applies to the context of social attitudes and behavior. A well-known example compared the effects of different ways of framing an anti-theft message to visitors at Arizona’s Petrified Forest National Park. Some visitors saw a sign that informed them that “[m]any past visitors have removed the petrified wood from the park,” others saw a sign reading “the vast majority of past visitors have left the petrified wood in the park,” and others only saw a message pleading with them to refrain from removing wood from the park (Cialdini et al. 2006). Those who were informed that the norm is to steal wood (“[m]any past visitors have removed the petrified wood from the park,”) stole more wood than people who received the other messages. The same principle is found in research on stereotyping. For example, Michelle Duguid and Melissa Thomas–Hunt (2015) found that telling people that “the vast majority of people have stereotypical preconceptions” leads them to express more stereotypes and to act in more stereotype-consistent ways, compared with people who are told that “very few people have stereotypical preconceptions” (Duguid & Thomas–Hunt 2015: 347).

Second, social practices influenced by norms typically involve rewards and punishments. This is a core tenet of the literature on the evolution of cooperation and has been explored extensively in game theoretic conditions, which often show that people are readily willing to incur costs for enforcing norms against norm-violators (see Henrich 2015 for review). Expectations of punishment also help to explain motivated reasoning in public policy contexts (see note #6) and public behavior, such as voting. Gerber and colleagues (2008) show that telling people that their participation in an election will be publicized is far more effective for increasing turnout than reminding people of their civic duty to vote or telling them that their participation will be monitored. In research on social attitudes and discrimination, one of the very founding concerns behind the “implicit revolution” was that people are motivated to conceal their biases and prejudices when they perceive those attitudes to be socially proscribed. The idea that rewards and punishment underlie the expression of bias is particularly central to neuroscientific approaches to prejudice and stereotyping (e.g., Amodio 2014).
De-biasing, skill, and intergroup virtue

Third, as Daniel Kelly and Taylor Davis (2018) argue, norms have intrinsic motivational force. That is, to acquire a norm typically entails feeling motivated to obey it. One feels compelled to tip around 20% in a restaurant once one knows that most people think that you ought to tip about 20% at a restaurant. This doesn’t mean that one will tip 20%, just that in the typical case one will feel some motivation to do so. One could say that norm perception has a tripartite structure. Acquiring a norm involves a belief about others’ beliefs or attitudes, a feeling that one ought to do what others do, and a behavioral inclination to act in a way consistent with the norm. Now, compare this to Jack Dovidio and colleagues’ characterization of the traditional conception of prejudice:

prejudice is typically conceptualized as an attitude that, like other attitudes, has a cognitive component (e.g., beliefs about a target group), an affective component (e.g., dislike), and a conative component (e.g., a behavioral predisposition to behave negatively toward the target group).

Dovidio et al. 2010: 524

If I am right that the expression of prejudice and stereotyping involves the psychology of norms, then there should be effective norms-based approaches to de-biasing. Paluck and colleagues (2016) offer a dramatic example. They sought to reduce student conflicts in 56 public middle schools (with a total of 24,191 students) by training small “seed groups” of students to model anti-conflict strategies. This modeling was explicitly intended to publicize new norms in the school—wearing wristbands signifying conflict-mitigation behavior, creating hashtag slogans, making posters linking seed group students’ images to anti-conflict slogans, etc.—not to persuade students about correct and incorrect behavior. Compared with control schools, disciplinary reports of student conflicts were reduced by 30% over one year. Even more strikingly, at schools in which at least 20% of the seed group was composed of students whom Paluck and colleagues identified as social referents, there was a 60% reduction in disciplinary reports. Social referents were defined as students whose behavior was likely to be observed by many other students. These students were identified with questionnaires aimed at mapping social networks, in particular by determining how many social connections each student at the schools had.

My contention is that Broockman and Kalla’s canvassers may have also been successful because they acted as social referents for the voters they met. One way this might have happened was by simply embodying the norm of acceptance of transgender people, thus shifting the voters’ perceptions of what other people think is normal and acceptable. Recall that the canvassers established a personal connection with the voters; this is important, given that people more readily uptake normative information from individuals with whom they identify (Festinger 1954; Wilder 1990). Moreover, there is evidence that people sometimes have more egalitarian ideals than they are willing to express publicly, when such ideals are perceived as socially unacceptable (e.g., Lessig (1995) argues that many American southerners privately opposed hiring discrimination against African-Americans prior to the civil rights movement, but only expressed support for anti-discrimination laws after legal changes in the 1960s changed their perceptions of what was socially (un)acceptable). Perhaps the canvassers demonstrated changing social norms, a demonstration that licensed the voters to express their egalitarian ideals.

Of course, my contention about the role of norm psychology in Broockman and Kalla’s research might be wrong. The contention could be tested in several ways in future research. For example, participants might be asked what they think people like them think about the
relevant laws. They might be asked whether they support penalties for people who do not share the view they hold. And their feelings about the laws themselves might be assessed; how motivated are they to vote? These three items follow from the features of norm psychology I emphasized above.

There are a number of open questions about the idea that skill in de-biasing involves becoming a social referent. Broockman and Kalla found no difference in the effectiveness of transgender canvassers and cisgender canvassers. But a norms-based approach suggests other possibly important traits to consider. For example, virtually all theories of norm psychology identify heuristics people use to choose from whom to learn important information. These “learning biases” (here “bias” is used in the non-pejorative sense) include prestige, success, and skill (i.e., people have biases toward learning from prestigious people, etc.). Does this suggest that de-biasers perceived as prestigious would be more effective than de-biasers who are perceived as ordinary? (The ghost of Stanley Milgram’s white-coated Yale-affiliated “scientist” lurks here, particularly given that subjects in a variant of Milgram’s original experiment were far less likely to shock the “learners” when the confederate instructing them to administer shocks identified himself as a lowly graduate student (Milgram 1975/2009).) Besides specific traits to consider, contextual variables are likely to influence a de-biase’s success in acting as a social referent. Tankard and Paluck (2016) argue, for example, that norms interventions are likely to be most useful in contexts where others’ beliefs and behavior aren’t publicly observable (e.g., whether others are likely to regularly check their tire pressure) and contexts where rewards and punishments are particularly strong (e.g., racial attitudes). Tankard and Paluck also address strategies de-biasers could study and utilize for successfully framing norms (e.g., describing changes in social attitudes, such as support for gay marriage, as gaining momentum, as in “more and more people are supporting gay marriage” (Tankard & Paluck 2016: 198).

### 39.5 Which norms?

Still, there is an open question about which norms de-biasing agents ought to represent. One reason for this is that there is surprisingly little empirical research focused on intergroup virtues rather than intergroup vices. Indeed, in The Nature of Prejudice, Allport wrote, “it is the pathology of bigotry and not the wholesome state of tolerance that, as a rule, interests social scientists” (Allport 1954: 425–426). Research in intergroup psychology has mostly followed this rule, focusing almost exclusively on vices such as hatred, fear, disgust, or indifference toward members of social groups. As a result, comparatively little is known about which norms de-biasing agents have and spread. The skills and techniques I’ve described thus far presume that intergroup virtue can be understood in terms of the absence of vices rather than the presence of virtues, such as compassion, open-mindedness, and respectfulness. Similarly, while philosophers have extensively considered intergroup prejudice, such as whether it is marked by ill will (Garcia 1999) or disrespect (Glasgow 2009), less attention has been paid to correlative virtues.

One exception is Lawrence Blum’s (2007) theory of racial virtue, which prioritizes feelings of comfort in intergroup interactions, recognition and respect of marginalized groups, and the capacity to see others as individuals. This makes intergroup virtue a complex construct consisting in several epistemic and practical virtues. Exemplars of intergroup virtue must have knowledge of common stereotypes and prejudices, and accept that they themselves are likely to hold biases. Related to this epistemic virtue is the ability to recognize when social categories such as race or gender are relevant (e.g., when someone is being discriminated against;
Madva 2016). This is a virtue of attention, of knowing when to pay attention to social categories and when not to. Exemplars of intergroup virtue must be motivated to act respectfully and display a degree of imperviousness to acquiring negative feelings about others in virtue of their social group membership. Practically, virtuous agents must know how to navigate complex intergroup interactions, exhibiting what has been called “interpersonal fluency” (Railton 2009, 2014; Anderson 2010; Brownstein & Madva 2012; Madva 2012). Moreover, being exemplary requires displaying each of these virtues over time and toward members of many different social groups.

If Blum is on the right track, the next question is how these virtues feed into the skills de-biasing agents possess. In what ways do the displays of these virtues affect others, if they do? In what contexts and conditions do they have salutary effects, and in what contexts and conditions do they perhaps backfire?

39.6 Conclusion

I have argued that skilled de-biasers resist the temptation to argue people out of their prejudices; treat intergroup contact as a goal; make personal connections with people using tools such as perspective-taking; utilize the psychology of norms to change people’s representations of what other people feel and believe; and, in some way that needs more definition, stand as exemplars of intergroup virtue and not just the absence of intergroup vice.

Does this all fall prey to the critique of “individual” approaches to prejudice-reduction? Are putative skilled de-biasers wasting their time trying to change individual minds, one-by-one? Should they not instead focus on laws, demography, and social practices? I won’t rehash the contours of the debate between individual and structural-institutional approaches to prejudice-reduction, but I will conclude by making two points about what I think the approach I’ve described here adds to that debate.

First, skill in de-biasing can have large societal effects, in part because of how the kinds of interventions I’ve described scale up. Recall Paluck and colleagues’ (2016) conflict-reduction intervention in middle schools. This touched the lives of approximately 12,000 students in 28 schools. It is not hard to imagine how it could grow far larger as well. Even Broockman and Kalla’s canvassers, who talked with voters one-on-one, spent only approximately ten minutes with each voter. That is certainly not an onerous investment, given the apparent payoff.

Second, the kinds of norms-based approaches I have described make an implicit concession to (some of) the structural-institutional critics. The concession is that changing social meanings, in Sally Haslanger’s (2015) sense—the social concepts, public narratives, and collective expectations with which we make meaning out of social situations—is indeed essential for creating social change. Crucially, changing people’s perceptions of norms is at the heart of this effort to change social meanings. Skilled de-biasers effectuate social change in part by representing new norms. They do not try to persuade people to embrace more inclusive social concepts, narratives, and expectations. Instead, they embody them.

Notes

1 To my knowledge, the original use of the term “de-biasing agents” is found in Kang and Banaji (2006).
3 See Beeghly (2015) for related discussion of descriptive versus evaluative senses of the term “stereotype.”
On defining prejudice, see McConahay and Hough (1976), Nosek and Banaji (2009), and Dixon et al. (2012). Note that in defining prejudice in terms of attitudes, I employ the psychologist’s sense of “attitudes,” not the philosopher’s (i.e., propositional attitudes, per se). On defining stereotypes, see Allport (1954), Ashmore and Del Boca (1981), and Stangor (2009).

See Madva and Brownstein (2016).

See, for example, Czopp et al. (2006).

See, for example, those employing “deep canvassing:” <https://knockeverydoor.org/how-it-works/index.html#deep-canvasing> and <www.ctctogether.org/>.

See Daniel Engber on Slate, for example: <https://slate.com/health-and-science/2018/01/weve-been-told-were-living-in-a-post-truth-age-dont-believe-it.html>.

For related critical discussion of Zack’s (2003) proposal regarding combatting racism by disseminating scientific information about the biology of race, see Kelly et al. (2010).

Engber (op. cit.) writes:

Yet even if boomerangs turn out to be unusual, [Brendan Nyhan] says, there’s little cause for optimism. Facts are, at best, “sometimes mildly effective” at displacing grabby lies, and corrections clearly aren’t working “if the standard is getting rid of misperceptions in the world.” Ullrich Ecker, the debunking expert who failed to reproduce Schwarz and Skurnik’s finding on the boomerang effect for facts and myths, agrees with Nyhan. “If there’s a strong motivation to hold on to a misconception, then often the corrections are ineffective. Whether or not they backfire, that’s up for debate,” he says. “But look, if it’s ineffective, that’s pretty much the same story as if there’s a small backfire effect.”

See also Swire et al. (2017).

See Madva and Brownstein (2016).

For example, see Brownstein et al. (2020) for application of this point to research on implicit bias.

The exception was for participants who lived with Asian American students. Their attitudes toward other outgroups became more negative.

But note that Towles-Schwen and Fazio (2006) found that randomly paired interracial freshmen roommates engaged in less joint activity and their relationships were more likely to dissolve compared with randomly paired white freshmen roommates.

See Kelly et al. (2010) for discussion of this point. See also Pettigrew and Tropp (2006), Pettigrew (2018), and Paluck et al. (2018) for discussion of the many contextual factors that contribute to the success—or lack of success—of intergroup contact as a tool for prejudice reduction.


Lest you confuse Broockman and Kalla’s study for the discredited and retracted LaCour and Green (2014), please note that Broockman and Kalla were widely credited for catching and publicizing LaCour and Green’s fraud (which was mostly LaCour’s fraud). Admirably, Broockman and Kalla then ran an actual study, which was similar in a number of ways to what LaCour and Green claimed to have done. It is also worth noting that Broockman and Kalla’s study is an exemplar in psychometric terms. It combines the use of random placebo control, a well-concealed outcome measure, longitudinal outcome measurement, and ecological validity in the field.

There was no effect of canvasser gender identity.

In some cases, canvassers would tell their own stories of feeling judged for being different.

See Mesoudi (2009) for discussion.

See also Blanchard et al. (1994) and Monteith et al. (1996) for similar findings.

An alternative explanation of this finding is that people are motivated by expected rewards for being known as a voter, rather than by expected punishment for being known as a non-voter.

See Madva and Brownstein (2016) for related discussion.


The full study included over 24,000 students in 56 schools, but half of these served as controls and received no intervention.
De-biasing, skill, and intergroup virtue

References


Michael Brownstein


—— (2002b) “Refocusing the question: Can there be skillful coping without propositional representations or brain representations?” *Phenomenology and the Cognitive Sciences*, 1, 413–425.


De-biasing, skill, and intergroup virtue


