CHAPTER 12
Individual and Structural Interventions
Alex Madva
ORCHID: /0000-0002-8222-5937

Abstract (not to be included in print version for metadata only): Given all that we have learned about bias and injustice, what can we do—and what should we do—to fight back? Chapter 12 introduces empirically-tested interventions for combating implicit (and explicit) bias and promoting a fairer world, from small daily-life debiasing tricks to larger structural interventions. Along the way, this chapter raises a range of moral, political, and strategic questions about these interventions, and stresses the importance of admitting that we don't yet have all the answers. We must encourage intellectual humility and dedicate ourselves to gathering as much knowledge as possible.

Changing the world is hard. Changing it for the better is usually harder than changing it for the worse. But why is positive change so difficult? Some answers are familiar. First, it's hard to get people to care, especially about problems in different places (physically or socially distant neighborhoods or countries) that don't confront us every day. Second, people are, understandably, wrapped up in pursuing personal goals (careers, families, hobbies). Third, and closely related to the first two reasons, many people feel like their votes and voices don't matter because the system is rigged (by corporate donors, gerrymandered voting districts, etc.) to make their political efforts pointless. Fourth, making change to promote *equality* is especially hard, because the haves are typically motivated to hold onto their advantages, and even to see their advantages as fair. Even the have-nots are easily hoodwinked into thinking that their *disadvantages* are fair when they're not (Jost, 2015). We derive comfort from believing we live in a merit-based society where, as long as you work hard enough, put your head down, and don't rile up political trouble, then your personal and professional life will go well. Fifth, people may

1

perceive that the world is so big, the problems so entrenched, and here I am, just a tiny, insignificant individual. How can I possibly make a difference?

This chapter helps to address such obstacles. But there's another reason why changing the world is so hard: we often don't know how to do it. After learning about implicit bias and related social ills, many people are persuaded that there is a problem, but they don't know what to do next. This chapter is meant to start chipping away at these knowledge gaps, to provide concrete tools to become less biased on an individual level, as well as to start thinking about potential larger-scale reforms for combatting bias, discrimination, and injustice—for promoting a fairer world.

But we shouldn't oversell what we know. Another aim of this chapter is to highlight remaining gaps in our knowledge, and encourage you, the reader, to do your part to fill them. True progress requires that we adopt an *experimental mindset*: test out different strategies and see how they go, then go back to the drawing board, revise our strategies, and test them again. By contrast, people concerned about racism, sexism, and other forms of bias and discrimination sometimes speak as if the changes we need to make are obvious. They're not. Consider two examples.

1.   Two Examples

Boxed Out

In the United States, we make it really hard for people with criminal records to get back, and stay on, their feet. One challenge is that many employers can, or sometimes must (by law), ask all applicants to "check a box" saying whether they have a criminal record. Asking ex-offenders to self-identify in job applications seems reasonable from the perspective of employers

(especially for certain jobs: maybe I don't want to hire someone convicted of decades of accounting fraud to be my accountant). But the result is that ex-offenders have a very hard time finding gainful employment, which in turn makes them more likely to become desperate to make ends meet, and then to re-offend and end up back in prison.

Some ex-offenders have it worse than others. In one field study in Milwaukee, the odds of getting a callback for an interview were 34% for white male applicants with no criminal record but only 17% for white men with a criminal record. Moreover, the odds were 14% for black male applicants without a record, but only 5% for black men with a record (Pager, 2003). The first finding to note is that white men *with* a record had slightly better odds than black men *without*. Two follow-up studies in New York City found much the same: white applicants who had literally just gotten out of prison were slightly more likely to get a callback or job offer than black and Latinx applicants with spotless records (Pager et al., 2009). These studies illustrate how devastating a criminal record can be for *anyone's* job prospects, but *especially* for people of color. 19 out of 20 ex-offenders of color can't make it past the initial screening, even just to land an *interview*, let alone secure the job.

This strikes many of us as unfair. Haven't these folks already paid their debt to society? But even if you don't find this unfair (maybe you think they "should have thought of that before committing the crime"), you might agree that the current system is *counterproductive* when it comes to reducing recidivism. We all share the goal of living safe from violence and crime, and many folks from both the left and right agree that much about the current criminal justice system makes ex-offenders of all races *more rather than less likely* to re-offend (Bibas, 2015).

So what should we do about it? While the causes and remedies are complex, one intuitive piece of the puzzle is to restrict employers' ability to demand that applicants "check the box."

Recent "Ban the Box" initiatives have prohibited employers from asking about applicants' criminal record until *after* they get past the initial screening process. As of April 2019, 35 states and over 150 cities have adopted some form of ban-the-box policy (Avery, 2019). It has been encouraging to see these steps being taken, but the results may not be what we were hoping for.

One study compared employment rates in regions before and after banning the box (Doleac and Hansen, 2016), and found that low-skilled black and Latinx men were marginally *less* likely to be employed after banning the box than before. It might be that banning the box reduces discrimination against ex-offenders overall but increases discrimination specifically against applicants of color with clean records. Why might this be? Maybe because when employers cannot ask upfront about criminal histories, they may (consciously or unconsciously) just *assume* that black and brown applicants have sketchy backgrounds. The researchers point to "a growing literature showing that well-intentioned policies that remove information about negative characteristics can do more harm than good" (9).

In response to such findings, what is an advocate for justice to do? Are we "damned if we do and damned if we don't?" Before drawing such pessimistic conclusions, we should take a step back and begin by admitting that we don't *already know* the best ways to address such problems. We should invest resources in *studying* the questions of how to facilitate ex-offender's reentry into public life, rather than patting ourselves on the back for making a small change that sounds good in the abstract but maybe does little to fix things, and might even make matters worse.

Anticipating these counterproductive results, Michelle Alexander wrote that, "banning the box is not enough. We must also get rid of the mind-set that puts black men 'in the box' (Alexander, 2012, p. 153; cf. Hernandez, 2017). It's not enough to change rules and policies; we must also consider *what's going on in the minds* of employers who consider applications from

ex-offenders, namely, stereotypes and prejudices about people of color. We must, collectively, overcome the mindsets (the feelings, assumptions, and implicit and explicit biases) that "box out" people of color from equal opportunities.

Leaving mothers behind

Upon first entering the workforce, men and women typically earn similar salaries. As their careers progress, that changes. Gradually, men's paychecks tend to grow significantly higher than women's. Men are also more likely to be promoted; in fact, women make up only 5% of the CEOs of the 500 largest US corporations (Zarya, 2018). One of the most significant factors here is parenthood. Women who become mothers tend to fall behind, whereas men who become fathers often do *even better* than men who don't become fathers (Aravena, 2015; Cudd, 2006, chap. 5; Miller, 2014; see also Lisa and Larry example in Ayala-López and Beeghly, Chapter 11, "Explaining Injustice: Structural Analysis, Bias, and Individuals"). (Parenthood is not the only factor behind the gender pay gap. Even if you compare single, childless men to single, childless women with comparable performances in comparable jobs, men still average about 5% higher salaries than women (e.g. Stewart and Valian, 2018, chap. 4). Also note that effects like the "fatherhood pay bonus" are strongest for high-skilled, cis straight white biological fathers married to biological mothers (Killewald, 2013; cf. Gasdaglis and Madva, Manuscript). There is some evidence for "breadwinner bonuses" and "caregiver penalties" in less gender-stereotypical parenting contexts (Bear and Glick, 2017), but a striking Norwegian study found that for *same-sex couples*, any pay gaps between parents due to having a child disappear within a few years (Andresen and Nix, 2019).)

What should we do about these gaps? Again, the causes and remedies here are complex. Much attention goes to family-friendly policies, such as allowing more flexible work schedules (Kliff, 2017), but especially to parental leave policies. Maybe if we give new parents more paid time off from work, then mothers will be less likely to fall behind. In fact, the US is the only developed country that doesn't guarantee paid leave for parents. Some employers opt to give mothers a few months of paid leave, but many don't. The US at least guarantees all parents 12 weeks of *unpaid* leave (although, of course, only parents who are not living paycheck-to-paycheck will be able to afford and take advantage of this policy).

As with Banning the Box, however, attempts to reform leave policies may have unintended negative consequences. To take a simple example, requiring employers to let women take maternity leave might make them less likely to hire women at all, or to promote them up the ranks. Consciously or unconsciously, perhaps they would rather not be on the hook for hiring or promoting a woman who (they assume) can just leave the job for months on end, when they could instead hire a man who (they assume) will continue working without significant interruption. One study looked at employment and promotion rates for women before and after the US passed its minimal policy guaranteeing unpaid leave (Thomas, 2016), finding that women overall became a little more likely to remain employed (i.e., to not be fired or quit), but a little less likely to get promotions, perhaps because employers were reluctant to make investments in early-career women who they feared wouldn't stick around. "The problem ends up being that all women, even those who do not anticipate having children or cutting back in hours, may be penalized," said Thomas in an interview (Miller, 2015).

Moreover, if mothers are being financially supported to take time off from work, but fathers are not, then fathers are going to stay on the job and have more time to move up the ranks

while new mothers lose experience, lose opportunities for promotion, and fall behind their male counterparts. The natural solution, then, would seem to be to encourage *paternity* leave in addition to maternity leave. Accordingly, Iceland includes 13 weeks of paid leave specifically for the non-childbearing parent, and most fathers take advantage of it (Kliff, 2018a). (Note that this policy is also more inclusive for gender-nonconforming parents.) Due to this and other aggressive efforts, Iceland has one of the lowest gender pay gaps in the world. They haven't eliminated the gap completely (still hovering around 5%!), partly because mothers tend to take *more* leave than fathers (Bershidsky, 2018).

But even completely gender-neutral parental benefits might not solve the problem. Everything depends on what parents *do* with the time set aside for caregiving. One study looked specifically at how certain gender-neutral parenting policies affected the odds of getting tenure for professors at 50 top-tier economics departments (Antecol et al., 2018). They found that some policies actually *increased* fathers' advantages over mothers, evidently because new fathers used the extra time to work on their research and strengthen their case for tenure, whereas mothers actually used (needed) the time to recover, breastfeed, and parent.

You can lead a horse to water but you can't make it drink. Maybe you can lead a father to the home but you can't make him parent. Or, much as Alexander said about ban-the-box campaigns, parental leave *policies* are not enough; we need to get rid of the mindsets that put one parent in the "breadwinner" box and the other in the "caregiver" box. We need to eliminate the conscious and unconscious expectations, habits, and preferences that treat fathers and mothers differently even when official laws and policies say they should be treated the same. Recall that whereas mothers are often paid less than childless women, fathers are typically paid *more* than childless men. One field study found that, even when their résumés were otherwise identical,

mothers were half as likely as childless women to be called back for an interview, whereas there was no penalty for fathers (Correll et al., 2007). Why might that be? Well, getting married and becoming a dad often strikes people as "responsible," sending the message that you'll be fully committed to your job because you need to support your family. By contrast, becoming a mom often sends the *opposite* message: employers think you are not fully committed to the job because you're going to take time off to take care of the kids. This is unfair nonsense. Most mothers work, and they are presumably just as committed to supporting their family as working fathers. And nearly two-thirds of fathers believe they should be spending more time with their children (Parker and Livingston, 2018).

There are also social pressures. Fathers who take full advantage of parental leave are sometimes seen as failing to live up to their stereotypical breadwinning role, whereas mothers who don't take advantage of all their parental leave are sometimes seen as failing to live up to their stereotypical caregiving role. In fact, over 25% of both men and women worldwide (and over 20% of men in North America) explicitly believe that women should stay home altogether (Gallup-International Labour Organization, 2017). On top of that, nearly half of American women think they should *both* work outside the home and maintain primary caregiving and housework responsibilities. (And remember that housework is *work*, even though it's often uncompensated and excluded from standard measures of economic output.)

Reforming parental leave policies won't—all by itself—dislodge all these biased beliefs and attitudes. Instead, dislodging biased attitudes may be essential to encouraging individual fathers to take full advantage of these policies and take on their fair share (i.e., half) of household responsibilities, *as well as* to expanding support for the sorts of powerful policy changes that Icelanders and others have explored. In other words, we need interventions specifically aimed at

changing individual hearts, habits, and minds, which may in turn be integral to bringing about necessary larger-scale social transformations and to ensuring that these transformations have the broadest and most durable impact on people's lives.

2.   Either/or versus Both/and

What examples like *Boxed Out* and *Leaving Mothers Behind* come to is an argument for insisting on the importance of individual-level *debiasing strategies*, which change individuals' biased assumptions, feelings, and habits, including their implicit and explicit social prejudices and stereotypes. These examples help to respond to an important criticism of the debiasing strategies we'll explore in this chapter, which is that they overlook more fundamental *structural* factors.

What are *structural factors*? The contrast is with *individual factors*. As Brownstein (Chapter 3, "Skepticism About Bias") and Ayala-López and Beeghly (Chapter 11, "Explaining Injustice: Structural Analysis, Bias, and Individuals") explain, different theorists draw the distinction between "individuals" and "structures" in different ways, but, to keep things simple here, we can think of structures as the *contexts* in which individuals operate. Individuals make choices, but they don't decide what the available options are. The range of available options, and how attractive or feasible each option is—that's part of the structure. Fathers in both Iceland and the US *can* choose to spend 13 weeks at home with their newborn. However, fathers in Iceland will be paid during that time, whereas fathers in the US will most likely not, and instead could be *fired* for being gone so long (past 12 weeks). Individuals have agency, the freedom to make choices, but it's the external context, or structure, that shapes which options are more or less available, feasible, and desirable.

Structures include all sorts of things, most obviously rules, procedures, and laws. They also include informal social norms and our physical and social environments. Communities of color are, for example, more likely to be exposed to environmental toxins than predominantly white communities (Ard, 2015). Consider Flint, Michigan. The water source for this predominantly black city was changed in 2014, which in turn exposed thousands of adults and children to lead poisoning (Clark, 2018). Being poisoned just by turning on the tap was a feature of their structure, part of the system in which they operated (and against which they protested, although authorities repeatedly insisted there was nothing to worry about). Other structural factors include the availability and affordability of high speed internet, public transportation, housing, childcare, and healthcare.

Roughly, structuralists argue that, when it comes to bringing about a more just society, we should deemphasize changing individual hearts and minds (think less about implicit bias, individual psychology, and ethics) and reemphasize changing structures (think more about sociology, urban planning, political philosophy, etc.). This means revising everything from social norms, to official laws, to the layout of physical space, to the broader systems governing how politicians are elected and how money and resources are extracted, generated, and distributed.

I am all in favor of overhauling existing social structures. In fact, basically everybody is! Disagreement revolves around about *which* structures to change, and how. Libertarians and anarchists want fewer laws and regulations; their opponents want more. Both want structural change, but they disagree about which. But let's grant structuralists the point that, for example, reforming family-friendly policies—like parental leave, work hour flexibility, and more affordable, accessible childcare—is necessary. We can also grant (just for the moment, for the

sake of argument) that these structural reforms will do more to promote fairness than combating individuals' stereotypes about gender, careers, and caregiving.

Does it follow that we should *prioritize* these structural reforms in general over changes in individual psychology in general? I don't see how (Madva, 2016). *Bringing about* these policy changes requires, at a minimum, changes in the beliefs, motivations, or actions of those individuals poised to help change policy. Such structural reforms are more likely if the relevant individuals are persuaded that the reforms are possible and desirable, and start acting to help bring the reforms about. Such reforms are more likely to "stick" and change behavior in enduring ways insofar as the individuals affected "buy into" them, or at least don't actively resist them (as in the case of employers and coworkers who think worse of fathers who take paternity leave). The most this sort of example can show, if it were correct, is that changing *certain* individual attitudes (in this case, stereotypes and personal preferences regarding gender, careers, and caregiving) is less relevant to bringing about necessary structural reform, and therefore to promoting fairness, than is changing *other* individual attitudes (e.g., changing individuals' motivation to reform parental leave policies!).

But note that for this example to make sense, we must also assume that individuals' attitudes about gender, careers, and caregiving exert no meaningful influence on their beliefs and motivations surrounding the reform of parental leave policies. That is an implausible assumption. Changing individuals' gender biases might very well be integral to drumming up support for reforming parental leave policies. Remember how many men and women alike still think mothers should stay at home or, even if working outside the home, still do the brunt of the homemaking? If views like these remain pervasive, how can we expect to generate enough enthusiasm and buy-in to make meaningful and durable changes to parenting policies? The

fundamental reason that it doesn't make sense to say things like, "don't worry about individuals' prejudices and stereotypes, just focus on changing structures," is that *individuals' prejudices and stereotypes are some of the most powerful factors shaping their willingness to support (or oppose) political and structural change* (Azevedo et al., 2019; Cooley et al., 2019; Harell et al., 2016; Monteith and Hildebrand, 2019; Mutz, 2018).

3.   More Lessons

With this in mind, the next sections recommend concrete strategies for combating implicit (and explicit) bias at both the individual and structural levels. But I want to make a few more points about *Boxed Out* and *Leaving Mothers Behind* first. What also becomes clear in these examples is that we don't yet know the best ways to address these problems. We need to study how to facilitate ex-offender's reentry into public life, and how to avoid penalizing mothers for working and fathers for caregiving. Too often we revise our policies with the expectation that they'll make a difference but then don't bother to check if they're helping or hurting (Dobbin et al., 2015). What's worse, even when policy changes don't make any positive difference, they can still give members of privileged groups the *false impression* that others are now being unfairly advantaged over them, leading them to become *more discriminatory* toward the disadvantaged than they were before (Dover et al., 2014; Kaiser et al., 2013).

So for starters, we need a healthy dose of *epistemic humility* when it comes to thinking about how to remedy social injustice (Medina, 2012; McHugh and Davidson, Chapter 9, "Epistemic Responsibility and Implicit Bias"). Humility means not overestimating how good you are at something, calibrating your level of confidence to your level of ability. Calling for *epistemic* humility, then, is a warning against overestimating how much we *know*, to be less arrogant and self-assured that the solutions to these problems are obvious.

The difficulty of fixing these problems also reveals how *multifaceted* they are. There is never a single law or stereotype or powerful individual solely to blame. We must figure out how the many puzzle pieces fit together, and then attack these problems in an accordingly multifaceted way. We must reject the forced choice: *either* pay attention to individuals *or* pay attention to structures. Resist that either/or framing and insist on a both/and framing. We should think about the complex ways that individuals fit into their structures, and how to change both individuals and structures *in tandem* to promote fairness (e.g., how to simultaneously undermine stereotypes *and* overturn discriminatory laws, norms, and built environments).

Relatedly, another lesson regards the importance of adopting an *experimental* approach to these issues. Since we can't know in advance what will work, we have to put interventions in place with every intention of *testing* them, and then going back to the drawing board if they don't work. What makes this so challenging is that collective motivation to change is usually a transient phenomenon: a crisis happens, then a bunch of changes are put in place, and then people forget about it—regardless whether the changes actually make things better or worse. Our initial plans have to include sub-plans to measure effects and re-assess. We need automatic checkpoints and triggers to determine whether our efforts are paying off.

Given our epistemic limitations and the multifaceted nature of these problems, when we try to make changes, we shouldn't put all our eggs in one basket. We should try out a bunch of things. We need what I call a *diversified experimentalism* (Madva, 2019a). If you're saving for retirement, investors say you shouldn't put it all in one company's stock, because that company might go bankrupt. Instead, you should diversify your investments across a bunch of companies. We should similarly diversify our experimental portfolio, and explore a bunch of different

individual and social experiments and interventions to see which ones stick and which ones stink.

With these lessons in mind, what follows are strategies that—according to one person in 2019—have enough evidential support to be worth a try. Future evidence might finetune, enrich, or even overturn the suggestions to follow. Here are some questions for you to think about for each strategy. First, most of these will work better in some contexts than others, and sometimes they might even backfire. Sorting out *when* and *when not* to use these tools is incredibly important—something for us to test in formal scientific settings as well as in the labs of lived experience. Ask yourself when these might be more useful, when they might be less useful, and when they might be downright ill-advised. Second, although these tools are framed in terms of what we as individuals can do, you can also consider how we might redesign our social institutions to encourage everybody to try them (a question we'll investigate further in §5, on structural reform). Third, ask yourself how these tools might be usefully *combined*. Perhaps we can couple different techniques together to make them more powerful (for more on the importance of mutually reinforcing strategies, see Madva, 2016, 2017, 2019b)?

4.   Six Debiasing Tools

Tool #1. The Life-Changing Magic of If-Then Plans

The first debiasing tool regards how to bridge the gap between intention and action. This tool is *if-then plans* (their official name is a mouthful: "implementation intentions"). These are concrete plans that specify when, where, or how we intend to put our broader aims and values into practice. The key idea is that we are more likely to follow through on our goals if we focus as

concretely as possible on the *contexts* for action and the specific thoughts and behaviors we will execute in those contexts. To get a sense of the difference between vague and concrete plans, contrast the first with the second plan in each of these two pairs:

1a. I'd like to cut back on smoking!

1b. If I feel a craving for cigarettes, then I will chew gum!

Or:

2a. My New Year's resolution is to work out more!

2b. When I leave work on Tuesdays, then I will go to the gym!

Which of these plans do you think will be more effective? 1a or 1b? What about 2a vs. 2b? Note that in each of these pairs, the first option just vaguely identifies the broad goal you're aiming for, but it doesn't say anything about how you intend to follow through on it. By contrast, the second option (step 1) identifies the specific contexts and obstacles that you're interested in and (step 2) highlights a concrete, straightforward plan about what to do in those contexts. Research suggests that if-then plans are easy to form (practice rehearsing them in your head a few times, or write them down), easy to remember, and easy to execute in the crunch. Meta-analyses consistently find that they can have dramatic effects on behavior and goal achievement (Gollwitzer and Sheeran, 2006; Toli et al., 2016).

If-then plans can be applied in pretty much any area of our lives where we recognize a gap between how we *think* we should act and how we actually *do* act. They are most studied in clinical settings related to healthy eating and substance use, but they also help combat implicit bias (Mendoza et al., 2010; Stewart and and Payne, 2008). For example, do you worry that you interrupt women more than men? Well, here's an if-then plan for you: "If she's talking, then I won't!" Do you sometimes suffer from stereotype threat or test anxiety? Before your next exam, mentally rehearse the following if-then plan, studied by Bayer and Gollwitzer (2007): "And if I start a new problem, then I will tell myself: I can solve it!"

But don't just take my word for it. It is incumbent upon us to formally and informally test these plans ourselves. Informally, that means trying these out in your daily life. See if they help. Whether they do or don't help, talk to someone you know about your personal experiment. Maybe find an if-then-planning buddy to share your intentions and experiences with. Or, if you yourself are a budding social scientist, then do a study on if-then plans! If you're a computer-coding entrepreneur, found a start-up company around an if-then planning app!

As you read through the tools to follow, think about examples of if-then plans that might help you put these tools into practice. You might also consider ways to "gamify" these tools. Are there fun games or apps that might be designed which can help us practice using these tools?

Tool #2: Approach Mindsets

The mindsets we take into our social interactions are essential for shaping how we get along, and whether we act in biased or unbiased ways. One study examined the different mindsets we might have when meeting or collaborating with someone from a different social group (e.g., a different

political party, religion, or, in this case, a different race). One group was given the goal to "avoid appearing prejudiced in any way during the interaction." This is an *avoidance* or *prevention-focused* mindset, where people focus on *what not to do*. People in this group had more difficult interactions and were more mentally drained after the conversation. Another group, however, was encouraged to adopt an *approach* or *promotion-focused* mindset: "approach the interaction as an opportunity to have an enjoyable intercultural dialogue" (Trawalter and Richeson, 2006, p. 409). People in this group found intergroup contact to be "rewarding rather than depleting" (411). Having an approach orientation makes conversations more likely to start off on the right foot and unfold in positive ways.

Approach mindsets have also been studied in clinical and habit-training contexts. For example, studies suggest that people with alcohol use disorder who (in conjunction with regular rehabilitative therapy) repeatedly practiced *approaching non-alcoholic drinks* and *avoiding alcoholic drinks* were less likely to relapse for at least one year (Eberl et al., 2013). Other studies find that simply *telling* participants that they are about to approach members of a different social group, or to approach healthy foods, has many of the same benefits as repeated practice, including reducing bias on the IAT (Van Dessel et al., 2017). In general, when we like something, we tend to approach it. These findings further suggest that when we practice or imagine approaching something, we also *come to like* it a bit more, too.

Tool #3: Common-Ground Mindsets

Every person has countless similarities and differences with everybody else, but which similarities and differences *count*? Which ones do we notice in day-to-day life? When people

17

meet a member of a noticeably different social group, they are more likely to look for and notice their differences than to pick up and build on what they share in common. Another potentially powerful mindset regards trying to find common ground when you meet a new person—even over something as trivial as whether you both prefer apples over oranges, or carpet over hardwood floors (Mallett et al., 2008). Are you both rooting for the same contestant on *The Bachelorette*, or against that team that seems to win the championship every year? Do you both see The Dress as black-and-blue or blue-and-gold? The effects of common-ground mindsets may be even stronger for more "self-revealing" questions, like the ones that come up in *Would You Rather?*. Would you rather a) be granted the answer to three questions or b) be granted the ability to resurrect one person (West et al., 2014)?

  With research like this in mind, Gehlbach and colleagues (2016; cf. Cortland et al., 2017; Robinson et al., 2019) had high school students and their teachers fill out a "get-to-know-you" survey with questions like, "If you could go to one sporting event, which of the following would you go to?" or "What do you do to de-stress?" Some teacher-student pairs learned what they had in common (e.g., maybe both would choose to go to the FIFA World Cup soccer finals, or maybe both de-stress by going for a walk). Compared to a control group that did not learn any shared-in-common facts, the intervention led to increased perceptions of similarity between instructors and students. It also boosted student achievement, especially for black and Latinx students who were traditionally underserved at the school. Strategies for promoting mindsets of common ground across group differences represent key avenues for future research. If you, dear reader, are a college instructor or student, maybe you can try this strategy out at your next meeting in office hours, or develop a common-ground icebreaker game.

Tool #4: The Power of Perspective

Part of what makes approach and common-ground mindsets effective is their ability to prompt *perspective-taking* across group boundaries. It's easier for members to understand, communicate, and collaborate on shared social and political projects when they are able to see things from each other's point of view. And it can be that much harder to take another's perspective when people differ in some obvious way, like geographical origin, religion, nationality, or (dis)ability. Face-to-face cooperation (see structural reform #4, below) remains the gold standard for promoting perspective-taking, but *narratives and games* are also useful here (see also structural reform #3). Community activists and scholars across the humanities and many social sciences have long emphasized the transformative power of narrative, and the empirical evidence bears them out. One study tested the effects of a twenty-minute, online "choose-your-own-adventure" game, in which Hungarians in their mid-20s occupied the perspective of an individual in the Hungarian Roma minority (Simonovits et al., 2018). Both immediately after the game and at least one month later, participants reported much less anti-Roma prejudice, as well as less prejudice toward another social group (refugees) who were not mentioned in the game. Participants were even 10% less likely to intend to vote for Hungary's far-right white-supremacist party.

Another study found that fictional, engaging narratives about intergroup contact and conflict can reduce explicit bias among young kids, high schoolers, and even undergraduates (in this case, students read passages about relations between Harry Potter's wizarding community and "Muggles," i.e., ordinary humans) (Vezzali et al., 2015). Perspective-taking interventions even reduce implicit bias, and lead, in turn, to more positive face-to-face interactions (Todd et al., 2011). Consider also that mock jurors encouraged to adopt the perspective of defendants

become less likely to find them guilty (Skorinko et al., 2014). See also McHugh and Davidson's discussions of "epistemic friction" and "world-traveling" (Chapter 7, "Epistemic Responsibility and Implicit Bias").

Tool #5: Persuasion and Value

But what does it actually mean to try to occupy another's perspective? Which part of others' perspectives should we try to occupy? Most perspective-taking interventions focus on imagining how other people *experience* the world (e.g., imagining what it's like to be in the Roma minority), but research suggests that people from different backgrounds also tend to emphasize different sorts of *ethical and political values* when they reason about what to do (Feinberg and Willer, 2015). This poses a problem for moral dialogue because each group tries to persuade the other in terms of the values they prioritize the most, rather than in terms of the values most salient to the person they're talking to. For example, left-leaning folks (e.g., typical "liberals" or "Democrats") tend to put a little more emphasis on fairness, reciprocity, and protecting the marginalized from harm, whereas right-leaning folks (e.g., "conservatives" or "Republicans") tend to place a bit more emphasis on patriotism, loyalty, and purity.

With this in mind, one useful task for enhancing perspective-taking and finding common ground may be to identify the moral values most central to the person you're talking to, and think about how to defend your goals in terms of those values. For example, a left-leaning person trying to persuade a right-leaning person to support marriage equality for everyone regardless of gender and sexual orientation might say, "Our fellow citizens of the United States of America deserve to stand alongside us, deserve to be able to make the same choices as everyone else

can… Our goal as Americans should be to strive for that ideal. We should lift our fellow citizens up, not bring them down" (Feinberg and Willer, 2015, p. 2). Right-leaning folks might better persuade left-leaning folks to support military spending by emphasizing its employment prospects for members of underemployed groups, including racial minorities, and explaining that "through the military, the disadvantaged can achieve equal standing and overcome the challenges of inequality and poverty" (p.7).

Some people (in particular, pristinely principled philosophers and other academics) find this strategy unbecoming. It can sound cynical and manipulative. Shouldn't we defend the right policies for the right reasons—the reasons we truly stand behind—rather than exploit rhetorical techniques that we don't actually agree with in order to get others to think and do what we want? This is a reasonable concern! There may also be unintended consequences of relying too heavily on strategies like this. For example, appealing to concerns about purity might be useful for someone trying to persuade others to care about the environment and healthcare ("Keep our lakes and rivers pure and unpolluted!... Keep our fellow citizens free from infection and disease!"), but problematic in other contexts. Purity has historically been at play in some of our worst racist impulses, from laws against "miscegenation" (interracial partnering) to the rhetorical strategies used to justify genocide, which portray the outgroup as an infestation of diseased, polluting insects that must be exterminated to preserve the ingroup's purity. Left-leaning folks may be understandably reluctant to let appeals to purity move back to the center of contemporary political discourse.

So if we try this strategy, we must do so carefully. But before rejecting it altogether, we should remember that making genuine headway toward improving intergroup communication and perspective-taking likely cannot just be about imagining others' experiences (e.g., imagining

"what it's like" for both police officers and black civilians to fear for their safety). It must surely also include a central role for taking each other's deeply felt values into account. Trying to identify the laws and policies that appeal to the widest range of people, because these laws can be defended in light of the broadest range of values, may be vital for bridging contemporary partisan divides.

Tool #6: Accentuate the Situation

As I write this, four of the American National Football League's (i.e., NFL's) thirty-two starting punters are from Australia (http://www.espn.com/nfl/players/_/position/p, 2018), as are many punters on American college teams (Bishara, 2018). Punters are highly specialized players who kick the ball both really far and with impressive accuracy to specific spots on the field. All the other punters are from the United States. Isn't it odd that Australians, who don't grow up playing American football, are making it in the NFL in this dedicated role? Why might this be? If we tried to explain this odd phenomenon by appealing to *internal* and *individual* factors, we might hypothesize that, since Australians are descended from British criminals, maybe they've inherited genes for physical strength or psychological grit. Maybe they have strong legs from outrunning the law or hopping after kangaroos? Or maybe not!

The real explanation is not about Australian players' genes or personality traits. It's about Australian players' *situations*, growing up in a country where "Aussie rules football," a game similar to rugby, is very popular. This sport also involves kicking long distances, in a similar way to American football (which itself grew out of rugby). Of course, the fact that this highly specific skill is valued in two distinct cultural contexts doesn't by itself explain why sure-footed

Aussies would abandon the game they love to learn a new one. What else is involved? How about the fact that Australians can make up to five times more money in the US and play professionally perhaps twice as long ("Australians in American football," 2018)? Once we consider their situations, there is nothing particularly *idiosyncratic* or *odd* about their choices. We don't have to appeal to any stereotypes about "Australian DNA" or "what Australian personalities are like" in order to explain their success in the NFL.

Too often we overlook situational factors and overplay individual factors when we try to explain things, especially when we are trying to explain the behavior of members of disadvantaged social groups. What's worse, the individual factors we appeal to are often stereotypes. For example, both men and women are more likely to explain a woman's anger in terms of her internal traits ("she is an angry person", "she is out of control"), but, when the very same reaction is exhibited by a man, people explain it in terms of his situation ("he is *justifiably* angry given the circumstances") (Brescoll and Uhlmann, 2008).

So when we try to understand others, we do well to consider whether we are explaining their behavior in terms of internal traits or situational forces (Levontin et al., 2013; Stewart et al., 2010). Why did Jamal show up late to work? Is it because he's lazy and doesn't want to work, or because that city-wide power outage last night stopped his alarm from going off? Maybe his brother's car broke down, which meant it fell to Jamal to drop his niece off at school. Why is Jamal acting uncomfortable now that he's arrived at work? Is it because he's got a challenging personality that doesn't "fit" our office culture, or because of anxiety that his predominantly white coworkers will assume that he came late just because he's black? Maybe he's uncomfortable because he rode up the elevator with a white woman who nervously clutched her purse as soon as he stepped in. With practice, we can make headway toward shifting our default

23

orientation away from stigmatizing, internalizing explanations and toward accentuating the situation.

5.   Structural Change

Speaking of redirecting our attention *away* from the idiosyncratic beliefs, feelings, and character traits that explain people's decisions and *toward* the broader situations that frame these decisions, this chapter now turns to concrete strategies for transforming our social institutions to combat bias and promote fairness. These strategies are roughly ordered from less to more transformational and impactful—and, accordingly, from less to more controversial.

Structural Reform #1. Decision-making criteria

All sorts of decisions can be influenced by implicit bias: Which candidate should I vote for? Is this defendant innocent or guilty? Which grade should I give this essay? How much should I tip my server? Should I swipe left or swipe right (Hutson et al., 2018)?

Yet our decisions are less biased when we make them on the basis of *clear criteria*. For example, one study asked participants to choose between two candidates for the job of chief of police (Uhlmann and Cohen, 2005). One candidate had extensive "street" experience (street smarts) but little formal education, and one had extensive formal education (book smarts) but little street experience. In addition, one candidate was a man, and one was a woman—and in different conditions they switched which was which. When the male candidate was street-smart and the female candidate was book-smart, participants reported that street smarts was the most important criterion for being chief of police, and recommended promoting the man. However,

another group had to choose between a street-smart woman and a book-smart man, and this group reported that *book smarts* was most important, and still recommended hiring the man. What's going on here? Participants had a gut feeling about who was the right fit for the job (chief of police = man), and then they combed through the résumés to find something to "justify" that initial gut feeling.

Nevertheless, this particular story has a happy ending. In a further condition, participants were asked to identify *in advance* which criterion (street experience or formal education) they valued most for the chief of police. When participants had settled on their decision-making criteria ahead of time, the bias in favor of hiring a man disappeared. Sometimes, preemptively settling on criteria—and sticking to them—is enough to eliminate the effects of bias.

Sometimes, but not always. Just because you've got some criteria doesn't necessarily mean they're any good! You might be *baking in* human bias or injustices without realizing it. This is a problem as organizations increasingly rely on computer algorithms to inform decisions, like whom to hire and whom to let out on parole (Johnson, n.d.; O'Neil, 2016). Many think that if a computer made the decision, it can't be biased, but everything depends on how we design these algorithms and what data we feed them. Algorithms, just like ordinary criteria, can make our decisions *better or worse* depending on how we use them.

We don't need high-tech examples to make the point. Consider law school admissions, which rely primarily on two criteria: LSAT performance and undergraduate GPA. Historically, LSAT scores have been given more weight in admissions (60%) compared to GPA (40%) (Crosby et al., 2003; Murphy et al., 2018). What's problematic about this weighting is that women have (on average) higher GPAs than men, but they also score (again, on average) worse on the LSAT. In other words, this weighting builds in an advantage for average male applicants

25

over average female applicants. This is all the more troubling when we consider that women's LSAT scores might be somewhat negatively affected by factors like stereotype threat and impostor syndrome (see Greene, Chapter 7, "Stereotype Threat, Identity, and the Disruption of Habit"), and given the evidence that *neither* the LSAT nor undergrad GPA actually predicts future bar-exam performance (that is, neither criterion correlates well with the exam that ultimately determines who becomes a practicing attorney).

Examples like this demonstrate the importance of *revisiting and revising* our criteria. (Maintain an experimental mindset!) One of the great benefits of criteria is how much easier they make it to collect and analyze data. When we make all our decisions based on gut feelings, it's very hard to tell if or where are decisions are going wrong. Once we start relying on clear criteria (LSATs, GPA, etc.), we can see whether these criteria are actually helping us make the best decisions, or are unfairly stacking the deck for some groups over others. Then we can continue to tweak the criteria to make sure they're treating everyone fairly and delivering the most accurate outcomes. One study of a large company revealed that women, people of color, and immigrant employees were being awarded smaller raises than white American men despite earning equivalent performance scores (i.e., despite meeting the same on-the-job criteria) (Castilla, 2008). In this case, their decision-making criteria seemed OK but managers weren't acting on them properly. Only by collecting data were they able to reveal this problem—after which point the company reformed its practices and has now all but eliminated disparities in raises (Castilla, 2015).

Part of what enabled this company to reform its practices was that working with clear decision-making criteria allows greater *transparency* for everybody involved about the reasons and procedures behind our decisions. Similarly, when teachers use a clear rubric for grading

essays, it's easier to be more open and transparent with students about the reasons for their grades. Clear rubrics can also point students (or applicants for jobs, promotions, grants, etc.) toward the precise areas where there's most room for improvement. Criteria also promote *accountability*. It is much easier to hold ourselves and others accountable for our decisions when we can justify them by reference to reasonable, mutually-agreed-upon standards.

Another virtue of criteria-based decision-making and data-collecting is that this strategy is relatively uncontroversial. People who are skeptical of ongoing efforts to combat bias and discrimination, or who are concerned about over-correcting and think the world is becoming unfair to white men, might nevertheless be on board for working together to settle on decision-making criteria. The goal here is to focus our attention only on the factors we deem most relevant for success, and bracket everything we think is irrelevant. That said, some criteria are more controversial than others. Which standards should universities use for admitting students, or for hiring professors? Should they only look at students' GPA and standardized test scores, or should they also consider, say, an applicant's demonstrated ability to overcome difficult circumstances, or to enrich the range of social perspectives on campus? Should hiring professors just be about easily measurable things like how many papers they publish or how often their papers get cited by others? Or should it count as a "plus" if, say, an instructor has a record of success in mentoring students who are the first generation in their family to go to college? When we are hiring a police chief, *is* book smarts or street smarts more important, or are they equally important? There are countless debates to have about these questions. A further benefit to having criteria is that they allow us to focus our discussions on *which* criteria we should use for making the best choice, and then to test out the criteria to see if they work.

Structural Reform #2. Anonymous Review

Another relatively uncontroversial tactic—which I swear by—is evaluating materials anonymously (see also Saul, 2013). Literally every time I grade papers, I am surprised, in two ways: first, by some of the outspoken students who are good at seeming smart in class but who evidently put less effort into their essays; but second, by the quieter students who blow me away with clear and insightful work. Simply put, the fact that I am surprised means that I'm biased. It means I have expectations. And those expectations can affect my grades without my noticing (Alesina et al., 2018; Carlana, 2019; Forgas, 2011; Harvey et al., 2016; van den Bergh et al., 2010). It's also, frankly, a relief to not have to think about these issues as I'm grading. This is speculative, but I think tamping down the salience of the author's identity sometimes frees up my mind to just dig in and focus on the paper itself.

Academic journals and professional conferences are increasingly moving toward anonymous review, and some early results are promising. Anonymous review seems especially well-suited to reducing *prestige bias*, when people assume, for example, that a paper must be high quality because it was written by an Ivy League professor. Another potential benefit of anonymous review is its power to *broadcast a commitment to fairness* to all parties involved. Some studies find that white women and people of color become more likely to submit their work to top-tier journals after the journals have transitioned to anonymous review, partly because they come to have more trust that their work will be evaluated fairly and accurately, without bias (Stewart and Valian, 2018, pp. 391–397).

Obviously, anonymous review is not always possible (see also Dominguez, Chapter 8, "Moral Responsibility for Implicit Biases") At some point, you may have to meet the person

you're evaluating face-to-face, so anonymity goes out the window. Even so, you may be able to incorporate anonymous review partially, for example, by reviewing some application materials anonymously (using clear criteria!) before the interview stage.

Where it's workable, anonymous review should again be a relatively uncontroversial tactic for those who are otherwise skeptical of efforts to address bias. Here we are not talking about giving any "extra consideration" to members of disadvantaged groups; we are just talking about how to maximize the chances that each individual essay or application is evaluated on its merits alone, rather than on less relevant factors like the prestige of their school. People who think they are already objective decision-makers might find anonymous review to be a nuisance, but they are less likely to protest that it's unjust.

Anonymous review may actually be more controversial among those already dedicated to resisting bias and injustice. This is because there are many facts that some people are better positioned to know than others (McHugh and Davidson, Chapter 7, "Epistemic Responsibility and Implicit Bias"), which means that sometimes we *should* take people's social identities into account when evaluating their claims (Alcoff, 2006). For example, if I want to know whether English is an easy language to learn, it would make more sense for me to ask a non-native English speaker than a native English speaker. People who had to learn English as a second language will know more about this topic than others. It would also make more sense for me to ask a linguist who has dedicated their career to studying second-language learning than it would to ask, say, an astrophysicist, brain surgeon, or four-star military general who, although maybe very smart, has not specialized in this area.

Rebecca Kukla applies this idea to the journal review process, arguing that, "knowing who wrote a piece is often important to assessing the value and meaning of what it says" (Kukla,

2018). Moreover, Kukla suspects that anonymous review is an essentially "conservative" policy that privileges people who write in a "mainstream" and "conventional voice," and may disadvantage people with less conventional styles or more radical ideas. Kukla further suggests that reviewers will often "be able to tell or at least make a strong guess about [the author's] identity. So the idea that anonymous review levels differences and removes biases is a myth." Just as appeals to *racial colorblindness* sometimes entrench whites' advantages over members of other races (Alexander, 2012), sometimes appeals to *total ignorance of identity* might further marginalize those outside mainstream identities or styles.

While anonymous review is certainly not a cure-all, a few points might be made in response to Kukla. First, we may misjudge how easy it is to guess an author's identity (Goues et al., 2018), perhaps because we are more likely to remember our correct guesses than our incorrect guesses, and also to forget all the times we couldn't hazard a guess at all. There are many contexts in which we are prone to overestimate our accuracy and abilities (Pronin et al., 2002), and this might be one of them. Consider a similar but much higher-stakes context: one of the leading causes of wrongful criminal conviction is eyewitness misidentification (The National Registry of Exonerations, 2019). Many witnesses who *think* they can identify the offender turn out to be wrong when the DNA evidence comes in. Second, perhaps there are ways to incorporate Kukla's valuable points about identity and knowledge without completely sacrificing anonymity. For example, an author might mention an aspect of their social identity when it's relevant to their argument (e.g., "As an able-bodied cis white man, I think I am more likely to be believed by police officers…") without completely blowing their cover and announcing their name. Although partial self-identification like this would open the door for some biases, it might nevertheless inhibit others (like prestige bias). Lastly, Kukla does not actually defend total de-

anonymity; she argues that attention to identity should play a role only at certain specific moments in the review process. Defenders of anonymous review should be open to the possibility of mixing anonymized and de-anonymized stages in our decision-making procedures—with the caveat that precisely when and how to practice anonymous review is an open-ended, empirical question that we cannot conclusively settle on the basis of anecdotal experience alone.

Social categories and hierarchy: Ingroup vs. outgroup, high status vs. low

Decision-making criteria and anonymous review are structural changes for blocking the influence of bias. For other promising examples of structural changes along these lines, see Stewart and Valian (2018, chaps. 5–10) and Glaser (2014, chap. 8). They are vital first steps, and they can be defended on narrow merit-based and colorblind principles that resonate with people from a variety of social and political backgrounds. Putting them into place requires some upfront costs of time and resources, but once they're in place, sticking to them is relatively easy. That said, their overall impact on bringing about a more just and less biased society is unknown. Note in particular that both strategies aim to *work around* our biases, leaving our preexisting prejudices and stereotypes in place but limiting their influence on important decisions. More substantial structural changes might aim to reduce or eliminate our biases altogether, or stop them from forming in the first place. With that in mind, we turn next to more impactful (and, correspondingly, more difficult) reforms. These will inevitably be more controversial, in some cases because they are explicitly color- or identity-conscious (which makes them more likely to trigger complaints of "reverse" discrimination against historically privileged groups), or because

they raise worries about objectionable forms of top-down "social engineering," where elitist administrators boss us around and try to remold our hearts and minds.

Figuring out how to uproot our biases requires understanding where they come from. Research on child development suggests that two of the most powerful factors behind the formation of implicit biases are ingroup-outgroup distinctions and status preferences (Dunham et al., 2008). From a very young age and continuing into adulthood (Axt et al., 2014), people tend to implicitly prefer ingroup members over outgroup members, and high-status group members over low-status group members (status is determined by factors like the group's average wealth, power, and visibility in celebrated jobs and positions). Since white people occupy a higher-status racial group, most whites implicitly prefer member of the ingroup over other races. Things are more complex for members of other racial groups, such as African Americans: for some, ingroup favoritism dominates social status (so they implicitly prefer blacks over whites); for others, status dominates ingroup favoritism (so they implicitly prefer whites over blacks); and many have no implicit preference either way (almost as if their ingroup favoritism and status biases cancel each other out).

It stands to reason that reforms for combatting implicit bias should be oriented toward disrupting these two major causes, by shifting widespread perceptions of the boundaries between "us" and "them," and by supporting egalitarian reforms that minimize status differences.

Structural Reform #3. Environmental Signals, Counterstereotypes and Representation

One powerful set of tools for breaking down us-vs.-them dichotomies and status hierarchies resides in the cues and signals in our environments (Murphy et al., 2018; Vuletich and Payne, 2019). What messages do our physical, social, and virtual spaces send about who does and

doesn't belong, and who will and won't excel? Even subtle cues about belonging and identity can have powerful effects on our biases, behaviors, and sense of identity. Recall a set of studies discussed by Nathifa Greene (Chapter 7, "Stereotype Threat, Identity, and the Disruption of Habit"). Students in high school or college were first brought into computer-science classrooms that were either filled with objects associated with science fiction and video games (*Star Trek* figurines and *World of Warcraft* posters) or decorated in a neutral fashion. Then the students answered a series of questions. Researchers found that being in the "geeky" classrooms dramatically reduced women's interest and expected success in computer science, but had no effect either way on men. In fact, girls and women are up to three times more likely to express interest in computer science in the neutral room (Master et al., 2016). Studies like this highlight the power of situations and environments to "influence students' sense of ambient belonging… or feeling of fit in an environment" (Cheryan et al., 2011, p. 1826). What messages are our environments sending about who "fits" in our neighborhoods, campuses, classrooms, offices, and dorm hallways?

Consider also a study on how attending an all-women's college affected undergraduate women's implicit biases regarding gender and leadership qualities (Dasgupta and Asgari, 2004). Beforehand, participants were quicker to implicitly associate names like "Emily" with traits stereotypical of women leaders, like "nurturing," but to associate names like "Greg" with stereotypically masculine traits, like "assertive." After one year, these implicit biases vanished. The same study also found that attending a coed school had the opposite effect on undergraduate women. After a year at the coed school, most women's implicit gender-leadership biases grew even more pronounced.

You might think the difference-maker was a more supportive atmosphere at the all-women's school. That's not what researchers found. Evidently, the difference boiled down to the total number of classes that students had with women math and science professors, because there was a larger pool of women math and science professors at the all-women's school. A closer look at the data showed that this was true *regardless* of which institution they attended. Women at the coed school who also managed to take a few math and science classes with women instructors also showed reduced implicit biases. Subsequent studies have consistently demonstrated that having a few role models "like you" can be very effective for increasing motivation, success, and belonging, and shaping implicit and explicit biases and goals (for a review, see Dasgupta, 2013). These studies reveal the debiasing power of *counterstereotypes*, exemplars in our social environments who buck our biased expectations.

More broadly, consider audience's profound reactions to blockbusters like *Wonder Woman*, *Black Panther*, and *Crazy Rich Asians* in 2017 and 2018, which prominently portrayed white women and people of color in a diverse range of counterstereotypical roles. Experimental and correlational studies demonstrate that exposure to outgroup members through engaging narratives can reduce bias, such as the effect of the sitcom *Will & Grace* on heterosexism (Schiappa et al., 2006; for a review, see Murrar and Brauer, 2019). In fact, an examination of long-term, population-level attitudes found that, from 2007 to 2016, the largest reduction in both implicit and explicit biases regarded biases against sexual minorities (Charlesworth and Banaji, 2019). It is difficult to suppose that changes in media representation (including journalistic coverage of the marriage equality debate) did not contribute to this change. We should therefore demand more cultural, racial, and other kinds of variety in our media diets. We can also take it upon ourselves to *create* more original and counterstereotypical content. Become the writers,

artists, actors, journalists and video game designers who enrich our media environment, and expand our sense of what's possible.

Unfortunately, the effects of seeing counterstereotypes sometimes differ for members of historically underrepresented versus historically advantaged groups. Sometimes whites perceive the increased prominence of people of color as a sign that whites are unfairly *losing* status, or they conclude that, if some historically disadvantaged folks can now become billionaires, CEOs, or President of the United States, that means that people of color *in general* no longer face serious obstacles (Alexander, 2012, p. 244ff; Wilkins et al., 2017). Given these predictable kinds of backlash, we cannot expect changes in environmental cues or media representation to do all the work (Madva, 2016). The empirical case for counterstereotypical cues may be better grounded in their power to provide possibility-expanding role models for members of disadvantaged groups than in their capacity to debias the advantaged.

Structural Reform #4. Intergroup Cooperation

Several of the individual debiasing tools listed above (such as 2 through 4, on approach and common-ground mindsets, and perspective-taking) are closely tied to a much broader and time-tested strategy: positive intergroup contact and cooperation. The basic recipe for reducing both explicit and implicit forms of bias is to get people from different groups to work together toward a common goal. Perhaps the most famous case was the desegregation of the US military in World War II. Soldiers who served in racially diverse units showed much less racial bias at the end of the war than soldiers in segregated units. For discussion of these and other early findings, see Pettigrew (1998). More humdrum examples are diverse sports teams and first-year college roommates (Shook and Fazio, 2008). Evidence also suggests that racially diverse juries are both

more likely than racially homogeneous juries to consider an array of perspectives and to more accurately recall case facts and testimony (Sommers, 2006).

Bearing in mind the potential for backlash discussed in the previous section, a key point to emphasize is that simply thrusting people from different groups into shared cubicles, classrooms, neighborhoods, or nation-states is not enough. We need intergroup *cooperation*, not just intergroup *conversation*, and definitely not just intergroup *physical proximity* (Enos, 2017; Putnam, 2007). In some contexts, intergroup contact and proximity can heighten rather than dampen bias, for example, when members of different groups are competing against each other for scarce resources. Productive intergroup contact must be 1) *frequent*, 2) on terms of relatively *equal social status*, 3) organized around *cooperating toward a common goal*, and 4) sponsored by *authority figures* who enforce conditions 1-3 (Allport, 1979; Anderson, 2010; Pettigrew, 2018). Think of the role that good team coaches (authority figures) can play in holding their players accountable for having frequent cooperative contact and ensuring that all players are recognized and valued for their contributions (thereby promoting relatively equal status on the team).

The challenge for all of us, then, is to constantly consider context: how can we structure our own social institutions to foster fruitful intergroup cooperation? In this vein, I often daydream about two sorts of initiatives. The first is a campus-wide team-based competition (e.g., a massive event modeled on pub team trivia, or a kickball tournament, or Assassin, or…?) in which all the different school clubs participate, but teams are composed of a mix of members from different clubs. So, e.g., some Campus Democrats and Campus Republicans would be on the same team, as would various athletes, members of the model UN, Chess team, Black Student Union, Muslim Student Association, and Queer Student Alliance. The second initiative I

envision is a large-scale American "domestic exchange program," in which high school students across the country spend one semester at another school in a very different situation, traveling between the center of the country and the coast, between north and south, between urban and rural, and so on. What would the world look like if everyone residing in the United States made a more concerted effort to learn face-to-face how the other half—or at least a few of the many, many other halves—live?

Some argue that we should make intergroup cooperation a central organizing principle for society, and strive to create racially integrated schools, businesses, residential neighborhoods, and voting districts (Adams, 2006; Anderson, 2010, chap. 6). Doing so might promote intergroup cooperation as well as ensure that that members of otherwise marginalized groups have access to the same high-quality education, healthcare, and job opportunities as members of more advantaged groups. In this way, proactive integrationist strategies have the potential to tackle the two major causes of implicit bias mentioned earlier: expanding our sense of who's "inside our group" and limiting the influence of inherited and unjust inequalities in social status.

However, it's one thing when an individual freely chooses to seek out more opportunities for productive intergroup cooperation; it's another thing when schools, businesses, and governments enforce top-down policies to integrate their citizens or subordinates. Shouldn't individuals have *freedom of association*, e.g., the ability to freely choose where to live and whom to befriend? One prominent strategy for promoting neighborhood integration is for the government to distribute vouchers which individuals can then use toward a down payment on a house, but only for houses in socioeconomically diverse communities. Some object to this policy on the grounds that it unfairly restricts people's choices (they say it's *paternalistic*), and because it is a *conditional* rather than *unconditional* distribution of goods (Goetz, 2018; Shelby, 2016; cf.

Madva, 2019a). If we take seriously that folks who live in concentrated poverty, such as urban ghettos, are denied the same access to the educational, healthcare, and employment opportunities available to residents of wealthy suburbs, then just devoting our efforts toward distributing one-way tickets out of the ghetto might seem problematic. Does it, in effect, amount to telling these folks that their neighborhoods and schools are beyond saving, and getting access to decent resources and opportunities requires uprooting their households, leaving behind their social networks, and moving away from services that cater to their distinctive needs and preferences (such as religious centers or hair salons)? What about the ghetto residents who would rather stay put, and what about the hostility or discrimination that those who move might face from new neighbors who don't want them there? One of the perennial obstacles to top-down initiatives to promote intergroup contact is the resistance of the advantaged, such as through "white flight." Historically, a prominent strategy for racially integrating schools in residentially segregated areas was to bus students across town. When these busing policies were put in place, many wealthy white parents just sent their kids to private school rather than let them share classrooms with less privileged children of color. (This trend was as common in Boston, Massachusetts as in Birmingham, Alabama.) Productive intergroup cooperation is a powerful intervention, but creating conditions for it that are both *effective in durable ways* and *fair to everyone involved* has proven difficult. Some conclude that we should think less about integrationist strategies, which aim to move people to resources, and focus more on redistributive strategies, which move resources—and power—to the people (Young, 2000).

Structural Reforms #5. Powers to the Peoples

Thus a final set of structural reforms tries to "cut to the chase" and focus on directly alleviating status hierarchies and inequalities rather than on breaking down psychological "us" vs. "them" dichotomies. These proposals are so varied that it is a disservice to lump them all together, but it is nevertheless worthwhile to draw readers' attention to a wider range of questions about how best to overhaul society.

First, consider ways to make the *distribution of goods* fairer. Some overhauls that might help to mitigate status differentials in society are universalistic and colorblind: guarantee everyone free healthcare, a basic income (Matthews, 2017), and better financial support for students all the way through college and grad school (Miller and Flores, 2016). Such proposals are seen as radically left-wing in the United States of 2019, but some are common and uncontroversial in other parts of the world. (For discussion of these and other proposals that would apply to everyone but likely have the most positive effects on the most oppressed, see (Movement for Black Lives, n.d.).) Others argue that more targeted, color-conscious redistributions to members of disadvantaged groups are necessary, such as reparations for past injustices. For an accessible introduction and back-and-forth about the virtues and vices of taking racial reparations seriously—not just for slavery but for the many injustices that have continued since, see Coates (2014a), Williamson (2014), and Coates (2014b).

A different set of tools for combating status disparities has less to do with making sure that goods like healthcare and education are equally available to all, and more to do with the well-functioning of democracy—in overtly political contexts but also more broadly in workplaces and classrooms. Note, for example, that most workplaces have a strong *hierarchical* structure, a top boss with nearly unlimited power to hire and fire whom they please, to force

39

employees to work long hours in unpleasant conditions, and so on. In response, some call for movement toward greater *workplace democracy* (Anderson, 2017; Frega et al., 2019). The idea is not for workers to vote democratically on every major decision. Instead, they might, for example, elect someone to represent their interests on the board that runs the company, or they might all jointly participate in deciding on the right decision-making criteria for getting promoted (structural reform #1). Defenders of these reforms argue that democracy is not just about "one person, one vote," but about a culture of equals actively participating and sharing in the governing and decision-making processes of social institutions, including businesses, schools, and even families. Such reforms do not eliminate the distinction between leader and led, but they might reduce the perception that those in leadership positions are inherently "better than" while others are "less than." That is, under these conditions, differences in leadership status would not entail differences in *esteem* or *respect*. Moreover, as mentioned in the previous section, getting people to cooperate on terms of *social equality* is key for reducing prejudice.

Perhaps a similar lesson should be applied in disadvantaged communities. Malcolm X (1963) famously contrasted "racial segregation" from "racial separation" in terms of community control. He argued that America's segregationist hierarchy was wrong not because it meant that whites and blacks lived separately (he disagreed that "separate is inherently unequal"), but because whites had an inordinate amount of *control* over black communities. His defense of "separation" called for more local control within predominantly black communities. The recent Movement for Black Lives has similarly called for more community control. Perhaps we should commit to redistributing resources to oppressed communities, but make sure that members of those communities have the democratic freedom to collectively decide for themselves how they'd like to use those resources (Goetz, 2018; Shelby, 2016). This would again be about

40

ensuring that residents of disadvantaged communities have the same status and respect as everyone else.

While we're at it, we might as well highlight broader reforms to promote well-functioning democracy, such as, in the US, the elimination of the electoral college (Prokop, 2016), reforming how political campaigns are financed (Kliff, 2018b), complex revisions to elections like ranked-choice voting (Nilsen, 2018), or redrawing political districts to prevent one side from holding onto power even when they lose the popular vote (Druke, 2017). Of course, for many of these reforms, the more impactful they'll be, the harder they'll be to set in motion and sustain. Returning more power to more people means taking power away from the elite few, and they don't want that to happen. In fact, even among members of disadvantaged groups, it's often difficult to appreciate just how unfair the current system is, let alone to mobilize against it (Jost, 2015).

When it comes to more dramatic efforts to reduce power inequalities between historically advantaged and disadvantaged groups, color-conscious proposals like reparations or integration often face more backlash than universalistic and colorblind policies. But it is easy to overstate the disparities in divisiveness between colorblind and color-conscious proposals. Evidence suggests that many universalistic reforms, such as expanding healthcare access for all citizens, are often *perceived* by whites as delivering unfair benefits to people of color (Maxwell and Shields, 2014). Even basic views about the reality of *climate change* are increasingly predicted by our attitudes about race (Benegal, 2018). Our politics have become so polarized that a whole range of superficially non-racial political and scientific issues are thoroughly intertwined with beliefs and biases about race (and about gender and other social categories!). As a result, we

should be skeptical about the possibility of identifying major structural transformations that somehow avoid divisive backlash altogether.

This returns me to a point I emphasized at the outset of this chapter related to *Boxed Out* and *Leaving Mothers Behind*. Some argue that, when it comes to bringing about a more just society, we should focus more on changing structures and less on changing hearts and minds (Mallon, 2018; Vuletich and Payne, 2019). But such claims repeatedly fail to appreciate the extent to which our hearts and minds prop up these unjust structures (Mandalaywala et al., 2018; Plaut, 2010; Ridgeway, 2014). Eliminating status differences between historically privileged and oppressed groups may reduce individuals' biases, but eliminating individuals' biases toward other social groups may increase support for measures to reduce status differences. More generally, if we want to change the status quo, we have to convince enough people that the status quo is unfair. So we have to examine the *psychological precursors* that lead people to get angry at injustice and animated to do something about it (Jost et al., 2017; van Zomeren, 2013). Structures change when attitudes change—and attitudes change when structures change… when attitudes change when structures change! Our beliefs, habits, biases and social structures are thoroughly interconnected and mutually reinforcing. Neither comes first; neither comes second; it must be both/and every step of the way.

6.   Conclusion

Leaders have a powerful role to play in shaping how individuals' interpret and respond to their own implicit biases. When authority figures send the message to their subordinates that their uncomfortable gut feelings are valid representations of social reality (e.g., telling straight people that the discomfort they feel toward gay people is justified), then vague implicit biases can transform into wholeheartedly endorsed explicit prejudices, and lead people to act in more

discriminatory ways (Cooley et al., 2015; Madva, 2019c). Leaders have the power to turn implicit biases into explicit prejudices and overt acts of discrimination. Such findings are disheartening when we think about people in leadership positions who stoke prejudice and division. But we can also respond to such findings as calls to action. The upshot is that *we need to become leaders*. Become the formal or informal leaders who demand change. Run for club president or political office, apply to move up the ranks at your job. When you get there, use your leadership status to broadcast your commitment to fairness and against bias. What signals are you sending by the policies you endorse, the opinions you respect, and the jokes you laugh at? How committed are you to fair, transparent, and shared decision-making and data-gathering, and to ensuring that everyone is treated with equal esteem and respect?

We don't just need leadership in politics, business, or education, i.e., the contexts where specific individuals are granted official supervisory status over others. Making headway will require a whole bunch of leaders tackling these complex problems from a whole bunch of angles, from many different positions in society. (Patrisse Cullors, co-founder of #BlackLivesMatter, calls for a movement that neither anoints a single figurehead leader nor becomes wholly leader*less*; the movement must instead be leader-full.) We need scientists to study causes and potential interventions, and we need activists and politicians to make changes. We also need *artists* to envision emancipatory alternatives. We need filmmakers and videogame designers and vloggers and science-fiction writers and journalists and therapists and lawyers and spiritual leaders. We need all the ingenuity, creativity, experimentation, and exploration we can get. The problem is so multifaceted, with so many different aspects to work on, that many people with many different strengths are necessary. Everyone can contribute, and contribute we must if meaningful, lasting change is going to come.

We must also not forget that *we're* the ones who have to make these changes happen. They will not happen on their own. Recall Martin Luther King Jr.'s warning about the dangers of believing that progress comes, slowly but inevitably, with the passage of time:

Such an attitude stems from a tragic misconception of time, from the strangely irrational notion that there is something in the very flow of time that will inevitably cure all ills. Actually, time itself is neutral; it can be used either destructively or constructively. More and more I feel that the people of ill will have used time much more effectively than have the people of good will. We will have to repent in this generation not merely for the hateful words and actions of the bad people but for the appalling silence of the good people. Human progress never rolls in on wheels of inevitability; it comes through the tireless efforts of men willing to be co workers with God, and without this hard work, time itself becomes an ally of the forces of social stagnation. We must use time creatively, in the knowledge that the time is always ripe to do right. (King Jr., 1963)

Recommendations for Further Reading

Please see published version for reading recommendations

Discussion questions

Please see published version for discussion questions

References

Adams, M., 2006. Radical Integration. Cal. L. Rev. 94, 261.
Alcoff, L.M., 2006. Visible Identities: Race, Gender, and the Self. Oxford University Press.

Alesina, A., Carlana, M., Ferrara, E.L., Pinotti, P., 2018. Revealing Stereotypes: Evidence from Immigrants in Schools (Working Paper No. 25333). National Bureau of Economic Research. https://doi.org/10.3386/w25333

Alexander, M., 2012. The New Jim Crow: Mass Incarceration in the Age of Colorblindness. The New Press, New York.

Allport, G.W., 1979. The Nature of Prejudice. Addison-Wesley Publishing Company.

Anderson, E., 2017. Private Government: How Employers Rule Our Lives (and Why We Don't Talk about It). Princeton University Press.

Anderson, E., 2010. The Imperative of Integration, Reprint edition. ed. Princeton University Press, Princeton, N.J.

Andresen, M.E., Nix, E., 2019. What Causes the Child Penalty? - Evidence from Same Sex Couples and Policy Reforms [WWW Document]. Statistics Norway, Discussion Papers No. 902. URL https://www.ssb.no/en/forskning/discussion-papers/what-causes-the-child-penalty (accessed 4.2.19).

Antecol, H., Bedard, K., Stearns, J., 2018. Equal but Inequitable: Who Benefits from Gender-Neutral Tenure Clock Stopping Policies? American Economic Review 108, 2420–2441. https://doi.org/10.1257/aer.20160613

Aravena, F., 2015. The Impact of Fatherhood on Men's Earnings in Canada. Université d'Ottawa / University of Ottawa. https://doi.org/10.20381/ruor-4017

Ard, K., 2015. Trends in exposure to industrial air toxins for different racial and socioeconomic groups: A spatial and temporal examination of environmental inequality in the U.S. from 1995 to 2004. Social Science Research 53, 375–390. https://doi.org/10.1016/j.ssresearch.2015.06.019

Australians in American football, 2018. . Wikipedia.

Avery, B., 2019. Ban the Box: U.S. Cities, Counties, and States Adopt Fair Hiring Policies. National Employment Law Project. URL https://www.nelp.org/publication/ban-the-box-fair-chance-hiring-state-and-local-guide/ (accessed 7.4.19).

Axt, J.R., Ebersole, C.R., Nosek, B.A., 2014. The Rules of Implicit Evaluation by Race, Religion, and Age. Psychol Sci 25, 1804–1815. https://doi.org/10.1177/0956797614543801

Azevedo, F., Jost, J.T., Rothmund, T., Sterling, J., 2019. Neoliberal Ideology and the Justification of Inequality in Capitalist Societies: Why Social and Economic Dimensions of Ideology Are Intertwined. Journal of Social Issues 75, 49–88. https://doi.org/10.1111/josi.12310

Bayer, U.C., Gollwitzer, P.M., 2007. Boosting scholastic test scores by willpower: The role of implementation intentions. Self and Identity 6, 1–19. https://doi.org/10.1080/15298860600662056

Bear, J.B., Glick, P., 2017. Breadwinner Bonus and Caregiver Penalty in Workplace Rewards for Men and Women. Social Psychological and Personality Science 8, 780–788. https://doi.org/10.1177/1948550616683016

Benegal, S.D., 2018. The spillover of race and racial attitudes into public opinion about climate change. Environmental Politics 27, 733–756. https://doi.org/10.1080/09644016.2018.1457287

Bershidsky, L., 2018. No, Iceland Hasn't Solved the Gender Pay Gap. Bloomberg.

Bibas, S., 2015. The Truth about Mass Incarceration. National Review.

Bishara, M., 2018. Why Australian punters are dominating American football. CNN.

Brescoll, V.L., Uhlmann, E.L., 2008. Can an angry woman get ahead? Status conferral, gender, and expression of emotion in the workplace. Psychol Sci 19, 268–275. https://doi.org/10.1111/j.1467-9280.2008.02079.x

Carlana, M., 2019. Implicit Stereotypes: Evidence from Teachers' Gender Bias. Q J Econ. https://doi.org/10.1093/qje/qjz008

Castilla, E.J., 2015. Accounting for the Gap: A Firm Study Manipulating Organizational Accountability and Transparency in Pay Decisions. Organization Science 26, 311–333. https://doi.org/10.1287/orsc.2014.0950

Castilla, E.J., 2008. Gender, Race, and Meritocracy in Organizational Careers. American Journal of Sociology 113, 1479–1526. https://doi.org/10.1086/588738

Charlesworth, T.E.S., Banaji, M.R., 2019. Patterns of Implicit and Explicit Attitudes: I. Long-Term Change and Stability From 2007 to 2016. Psychol Sci 0956797618813087. https://doi.org/10.1177/0956797618813087

Cheryan, S., Meltzoff, A.N., Kim, S., 2011. Classrooms matter: The design of virtual classrooms influences gender disparities in computer science classes. Computers & Education 57, 1825–1835. https://doi.org/10.1016/j.compedu.2011.02.004

Clark, A., 2018. 'Nothing to worry about. The water is fine': how Flint poisoned its people. The Guardian.

Coates, T.-N., 2014a. The Case for Reparations. The Atlantic.

Coates, T.-N., 2014b. The Case for American History. The Atlantic.

Cooley, E., Brown-Iannuzzi, J.L., Boudreau, C., 2019. Shifting Stereotypes of Welfare Recipients Can Reverse Racial Biases in Support for Wealth Redistribution. Social Psychological and Personality Science 194855061982906. https://doi.org/10.1177/1948550619829062

Cooley, E., Payne, B.K., Loersch, C., Lei, R., 2015. Who Owns Implicit Attitudes? Testing a Metacognitive Perspective. Personality and Social Psychology Bulletin 41, 103–115. https://doi.org/10.1177/0146167214559712

Correll, S.J., Benard, S., Paik, I., 2007. Getting a Job: Is There a Motherhood Penalty? American Journal of Sociology 112, 1297–1339. https://doi.org/10.1086/511799

Cortland, C.I., Craig, M.A., Shapiro, J.R., Richeson, J.A., Neel, R., Goldstein, N.J., 2017. Solidarity through shared disadvantage: Highlighting shared experiences of discrimination improves relations between stigmatized groups. Journal of Personality and Social Psychology 113, 547–567. https://doi.org/10.1037/pspi0000100

Crosby, F.J., Iyer, A., Clayton, S., Downing, R.A., 2003. Affirmative action. Psychological data and the policy debates. Am Psychol 58, 93–115.

Cudd, A.E., 2006. Analyzing Oppression, 1 edition. ed. Oxford University Press, New York.

Dasgupta, N., 2013. Implicit Attitudes and Beliefs Adapt to Situations: A Decade of Research on the Malleability of Implicit Prejudice, Stereotypes, and the Self-Concept. Advances in Experimental Social Psychology 47, 233–279.

Dasgupta, N., Asgari, S., 2004. Seeing Is Believing: Exposure to Counterstereotypic Women Leaders and its Effect on Automatic Gender Stereotyping. Journal of Experimental Social Psychology 40, 642–658.

Dobbin, F., Schrage, D., Kalev, A., 2015. Rage against the Iron Cage: The Varied Effects of Bureaucratic Personnel Reforms on Diversity. Am Sociol Rev 80, 1014–1044. https://doi.org/10.1177/0003122415596416

Doleac, J.L., Hansen, B., 2016. Does "Ban the Box" Help or Hurt Low-Skilled Workers? Statistical Discrimination and Employment Outcomes When Criminal Histories are Hidden (Working Paper No. 22469). National Bureau of Economic Research. https://doi.org/10.3386/w22469

Dover, T.L., Major, B., Kaiser, C.R., 2014. Diversity initiatives, status, and system-justifying beliefs: When and how diversity efforts de-legitimize discrimination claims. Group Processes & Intergroup Relations 17, 485–493. https://doi.org/10.1177/1368430213502560

Druke, G., 2017. The Gerrymandering Project. FiveThirtyEight.

Dunham, Y., Baron, A.S., Banaji, M.R., 2008. The development of implicit intergroup cognition. Trends in Cognitive Sciences 12, 248–253. https://doi.org/10.1016/j.tics.2008.04.006

Eberl, C., Wiers, R.W., Pawelczack, S., Rinck, M., Becker, E.S., Lindenmeyer, and J., 2013. Approach Bias Modification in Alcohol Dependence: Do Clinical Effects Replicate and for Whom Does It Work Best? Developmental Cognitive Neuroscience 4, 38–51.

Enos, R.D., 2017. The space between us: social geography and politics. Cambridge University Press.

Feinberg, M., Willer, R., 2015. From Gulf to Bridge: When Do Moral Arguments Facilitate Political Influence? Personality and Social Psychology Bulletin 41, 1665–1681. https://doi.org/10.1177/0146167215607842

Forgas, J.P., 2011. She just doesn't look like a philosopher…? Affective influences on the halo effect in impression formation. European Journal of Social Psychology 41, 812–817.

Frega, R., Herzog, L., Neuhäuser, C., 2019. Workplace democracy—The recent debate. Philosophy Compass. https://doi.org/10.1111/phc3.12574

Gallup-International Labour Organization, 2017. Towards a better future for women and work: Voices of women and men.

Gasdaglis, K., Madva, A., Manuscript. Intersectionality as a Regulative Ideal.

Gehlbach, H., Brinkworth, M.E., King, A.M., Hsu, L.M., McIntyre, J., Rogers, T., 2016. Creating birds of similar feathers: Leveraging similarity to improve teacher–student relationships and academic achievement. Journal of Educational Psychology 108, 342–352. https://doi.org/10.1037/edu0000042

Glaser, J., 2014. Suspect Race: Causes and Consequences of Racial Profiling. Oxford University Press.

Goetz, E.G., 2018. The One-Way Street of Integration: Fair Housing and the Pursuit of Racial Justice in American Cities. Cornell University Press.

Gollwitzer, P.M., Sheeran, P., 2006. Implementation Intentions and Goal Achievement: A Meta-analysis of Effects and Processes, in: Advances in Experimental Social Psychology. Elsevier, pp. 69–119.

Goues, C.L., Brun, Y., Apel, S., Berger, E., Khurshid, S., Smaragdakis, Y., 2018. Effectiveness of Anonymization in Double-blind Review. Communications of the ACM 61, 30–33. https://doi.org/10.1145/3208157

Harell, A., Soroka, S., Iyengar, S., 2016. Race, prejudice and attitudes toward redistribution: A comparative experimental approach. European Journal of Political Research 55, 723–744. https://doi.org/10.1111/1475-6765.12158

Harvey, K.E., Suizzo, M.-A., Jackson, K.M., 2016. Predicting the Grades of Low-Income–Ethnic-Minority Students From Teacher-Student Discrepancies in Reported Motivation.

The Journal of Experimental Education 84, 510–528.
https://doi.org/10.1080/00220973.2015.1054332

Hernandez, P., 2017. Ban-the-Box "Statistical Discrimination" Studies Draw the Wrong Conclusions. National Employment Law Project. URL https://www.nelp.org/blog/ban-the-box-statistical-discrimination-studies-draw-the-wrong-conclusions/ (accessed 10.21.18).

Hutson, J.A., Taft, J.G., Barocas, S., Levy, K., 2018. Debiasing Desire: Addressing Bias & Discrimination on Intimate Platforms. Proceedings of the ACM on Human-Computer Interaction 2, 1–18. https://doi.org/10.1145/3274342

Johnson, G., n.d. Algorithmic Bias: on the Implicit Biases of Social Technology.

Jost, J.T., 2015. Resistance to change: A social psychological perspective. Social research 82, 607–636.

Jost, J.T., Becker, J., Osborne, D., Badaan, V., 2017. Missing in (Collective) Action: Ideology, System Justification, and the Motivational Antecedents of Two Types of Protest Behavior. Current Directions in Psychological Science 26, 99–108. https://doi.org/10.1177/0963721417690633

Kaiser, C.R., Major, B., Jurcevic, I., Dover, T.L., Brady, L.M., Shapiro, and J.R., 2013. Presumed Fair: Ironic Effects of Organizational Diversity Structures. Journal of Personality and Social Psychology 104, 504–519.

Killewald, A., 2013. A Reconsideration of the Fatherhood Premium: Marriage, Coresidence, Biology, and Fathers' Wages. Am Sociol Rev 78, 96–116. https://doi.org/10.1177/0003122412469204

King Jr., M.L., 1963. Letter From a Birmingham Jail [WWW Document]. URL https://kinginstitute.stanford.edu/king-papers/documents/letter-birmingham-jail (accessed 1.6.19).

Kliff, S., 2018a. A stunning chart shows the true cause of the gender wage gap. Vox.

Kliff, S., 2018b. Seattle's radical plan to fight big money in politics. Vox.

Kliff, S., 2017. The truth about the gender wage gap. Vox.

Kukla, R., 2018. Diversity and Philosophy Journals: How to Avoid Conservative Gatekeeping. Blog of the APA. URL https://blog.apaonline.org/2018/08/30/diversity-and-philosophy-journals-how-to-avoid-conservative-gatekeeping/ (accessed 1.2.19).

Levontin, L., Halperin, E., Dweck, C.S., 2013. Implicit theories block negative attributions about a longstanding adversary: The case of Israelis and Arabs. Journal of Experimental Social Psychology 49, 670–675. https://doi.org/10.1016/j.jesp.2013.02.002

Madva, A., 2019a. Integration, Community, and the Medical Model of Social Injustice. Journal of Applied Philosophy 0. https://doi.org/10.1111/japp.12356

Madva, A., 2019b. The Inevitability of Aiming for Virtue, in: Sherman, B.R., Goguen, S. (Eds.), Overcoming Epistemic Injustice. Rowman and Littlefield.

Madva, A., 2019c. Social Psychology, Phenomenology, and the Indeterminate Content of Unreflective Racial Bias, in: Lee, E.S. (Ed.), Race as Phenomena: Between Phenomenology and Philosophy of Race. Rowman & Littlefield International, Lanham, pp. 87–106.

Madva, A., 2017. Biased against Debiasing: On the Role of (Institutionally Sponsored) Self-Transformation in the Struggle against Prejudice. Ergo, an Open Access Journal of Philosophy 4, 145–179. http://dx.doi.org/10.3998/ergo.12405314.0004.006

Madva, A., 2016. A Plea for Anti-Anti-Individualism: How Oversimple Psychology Misleads Social Policy. Ergo, an Open Access Journal of Philosophy 3, 701–728. https://doi.org/10.3998/ergo.12405314.0003.027

Mallett, R.K., Wilson, T.D., Gilbert, D.T., 2008. Expect the unexpected: Failure to anticipate similarities leads to an intergroup forecasting error. Journal of Personality and Social Psychology 94, 265–277. https://doi.org/10.1037/0022-3514.94.2.94.2.265

Mallon, R., 2018. Constructing race: racialization, causal effects, or both? Philosophical Studies 175, 1039–1056. https://doi.org/10.1007/s11098-018-1069-8

Mandalaywala, T.M., Amodio, D.M., Rhodes, M., 2018. Essentialism Promotes Racial Prejudice by Increasing Endorsement of Social Hierarchies. Social Psychological and Personality Science 9, 461–469. https://doi.org/10.1177/1948550617707020

Master, A., Cheryan, S., Meltzoff, A.N., 2016. Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. Journal of Educational Psychology 108, 424. https://doi.org/10.1037/edu0000061

Matthews, D., 2017. A basic income really could end poverty forever [WWW Document]. Vox. URL https://www.vox.com/policy-and-politics/2017/7/17/15364546/universal-basic-income-review-stern-murray-automation (accessed 1.8.19).

Maxwell, A., Shields, T., 2014. The Fate of Obamacare: Racial Resentment, Ethnocentrism and Attitudes about Healthcare Reform. Race Soc Probl 6, 293–304. https://doi.org/10.1007/s12552-014-9130-5

Medina, J., 2012. The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations, Studies in Feminist Philosophy. Oxford University Press, Oxford, New York.

Mendoza, S.A., Gollwitzer, Peter.M., Amodio, D.M., 2010. Reducing the Expression of Implicit Stereotypes: Reflexive Control Through Implementation Intentions. Personality and Social Psychology Bulletin 36, 512–523. https://doi.org/10.1177/0146167210362789

Miller, B., Flores, A., 2016. How lessons from health care and housing could fix higher education affordability [WWW Document]. Vox. URL https://www.vox.com/2016/5/26/11767096/higher-education-affordability (accessed 1.8.19).

Miller, C.C., 2015. When Family-Friendly Policies Backfire. The New York Times.

Miller, C.C., 2014. The Motherhood Penalty vs. the Fatherhood Bonus. The New York Times.

Monteith, M.J., Hildebrand, L.K., 2019. Sexism, perceived discrimination, and system justification in the 2016 U.S. presidential election context. Group Processes & Intergroup Relations 136843021982668. https://doi.org/10.1177/1368430219826683

Movement for Black Lives, n.d. Platform [WWW Document]. The Movement for Black Lives. URL https://policy.m4bl.org/platform/ (accessed 1.6.19).

Murphy, M.C., Kroeper, K.M., Ozier, E.M., 2018. Prejudiced Places: How Contexts Shape Inequality and How Policy Can Change Them. Policy Insights from the Behavioral and Brain Sciences 5, 66–74. https://doi.org/10.1177/2372732217748671

Murrar, S., Brauer, M., 2019. Overcoming Resistance to Change: Using Narratives to Create More Positive Intergroup Attitudes. Current Directions in Psychological Science.

Mutz, D.C., 2018. Status threat, not economic hardship, explains the 2016 presidential vote. PNAS 115, E4330–E4339. https://doi.org/10.1073/pnas.1718155115

Nilsen, E., 2018. Maine voters blew up their voting system and started from scratch. Vox.

O'Neil, C., 2016. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy, 1 edition. ed. Crown, New York.

Pager, D., 2003. The Mark of a Criminal Record. American Journal of Sociology 108, 937–975. https://doi.org/10.1086/374403

Pager, D., Western, B., Bonikowski, B., 2009. Discrimination in a Low-Wage Labor Market: A Field Experiment. Am Sociol Rev 74, 777–799. https://doi.org/10.1177/000312240907400505

Parker, K., Livingston, G., 2018. 7 facts about American fathers. Pew Research Center. URL http://www.pewresearch.org/fact-tank/2018/06/13/fathers-day-facts/ (accessed 1.7.19).

Pettigrew, T.F., 2018. The Emergence of Contextual Social Psychology. Pers Soc Psychol Bull 44, 963–971. https://doi.org/10.1177/0146167218756033

Pettigrew, T.F., 1998. Intergroup Contact Theory. Annual Review of Psychology 49, 65–85. https://doi.org/10.1146/annurev.psych.49.1.65

Plaut, V.C., 2010. Diversity Science: Why and How Difference Makes a Difference. Psychological Inquiry 21, 77–99. https://doi.org/10.1080/10478401003676501

Prokop, A., 2016. Why the Electoral College is the absolute worst, explained. Vox.

Pronin, E., Lin, D.Y., Ross, L., 2002. The Bias Blind Spot: Perceptions of Bias in Self Versus Others. Personality and Social Psychology Bulletin 28, 369–381. https://doi.org/10.1177/0146167202286008

Putnam, R.D., 2007. E Pluribus Unum: Diversity and Community in the Twenty-first Century-The 2006 Johan Skytte Prize Lecture. Scandinavian Political Studies 30, 137–174.

Ridgeway, C.L., 2014. Why Status Matters for Inequality. Am Sociol Rev 79, 1–16. https://doi.org/10.1177/0003122413515997

Robinson, C.D., Scott, W., Gottfried, M.A., 2019. Taking It to the Next Level: A Field Experiment to Improve Instructor-Student Relationships in College. AERA Open 5, 2332858419839707. https://doi.org/10.1177/2332858419839707

Saul, J., 2013. Implicit Bias, Stereotype Threat, and Women in Philosophy, in: Hutchinson, K., Jenkins, F. (Eds.), Women in Philosophy: What Needs to Change? Oxford University Press, pp. 39–60.

Schiappa, E., Gregg, P.B., Hawes, D.E., 2006. Can One TV Show Make a Difference?a Will & Grace and the Parasocial Contact Hypothesis. Journal of Homosexuality 51, 15–37. https://doi.org/10.1300/J082v51n04_02

Shelby, T., 2016. Dark ghettos: injustice, dissent, and reform. The Belknap Press of Harvard University Press, Cambridge, Massachusetts.

Shook, N.J., Fazio, R.H., 2008. Interracial Roommate Relationships: An Experimental Field Test of the Contact Hypothesis. Psychological Science 19, 717–723.

Simonovits, G., Kézdi, G., Kardos, P., 2018. Seeing the World Through the Other's Eye: An Online Intervention Reducing Ethnic Prejudice. American Political Science Review 112, 186–193. https://doi.org/10.1017/S0003055417000478

Skorinko, J.L., Laurent, S., Bountress, K., Nyein, K.P., Kuckuck, D., 2014. Effects of perspective taking on courtroom decisions. Journal of Applied Social Psychology 44, 303–318. https://doi.org/10.1111/jasp.12222

Sommers, S.R., 2006. On racial diversity and group decision making: identifying multiple effects of racial composition on jury deliberations. J Pers Soc Psychol 90, 597–612. https://doi.org/10.1037/0022-3514.90.4.597

Stewart, A.J., Valian, V., 2018. An Inclusive Academy: Achieving Diversity and Excellence. The MIT Press, Cambridge, MA.

Stewart, B.D., and Payne, B.K., 2008. Bringing Automatic Stereotyping under Control: Implementation Intentions as Efficient Means of Thought Control. Personality and Social Psychology Bulletin 34, 1332–1345.

Stewart, T.L., Latu, I.M., Kawakami, K., Myers, A.C., 2010. Consider the situation: Reducing automatic stereotyping through Situational Attribution Training. Journal of Experimental Social Psychology 46, 221–225. https://doi.org/10.1016/j.jesp.2009.09.004

The National Registry of Exonerations, 2019. Exonerations Contributing Factors By Crime [WWW Document]. URL https://www.law.umich.edu/special/exoneration/Pages/ExonerationsContribFactorsByCrime.aspx# (accessed 1.2.19).

Thomas, M., 2016. The Impact of Mandated Maternity Benefits on the Gender Differential in Promotions: Examining the Role of Adverse Selection. Institute for Compensation Studies.

Todd, A.R., Bodenhausen, G.V., Richeson, J.A., Galinsky, A.D., 2011. Perspective taking combats automatic expressions of racial bias. Journal of Personality and Social Psychology 100, 1027–1042. https://doi.org/10.1037/a0022308

Toli, A., Webb, T.L., Hardy, G.E., 2016. Does forming implementation intentions help people with mental health problems to achieve goals? A meta-analysis of experimental studies with clinical and analogue samples. British Journal of Clinical Psychology 55, 69–90. https://doi.org/10.1111/bjc.12086

Trawalter, S., Richeson, J.A., 2006. Regulatory focus and executive function after interracial interactions. Journal of Experimental Social Psychology 42, 406–412. https://doi.org/10.1016/j.jesp.2005.05.008

Uhlmann, E.L., Cohen, G.L., 2005. Constructed Criteria: Redefining Merit to Justify Discrimination. Psychological Science 16, 474–480.

van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., Holland, R.W., 2010. The Implicit Prejudiced Attitudes of Teachers: Relations to Teacher Expectations and the Ethnic Achievement Gap. American Educational Research Journal 47, 497–527.

Van Dessel, P., Gawronski, B., Smith, C.T., De Houwer, J., 2017. Mechanisms underlying approach-avoidance instruction effects on implicit evaluation: Results of a preregistered adversarial collaboration. Journal of Experimental Social Psychology 69, 23–32.

van Zomeren, M., 2013. Four Core Social-Psychological Motivations to Undertake Collective Action. Social and Personality Psychology Compass 7, 378–388.

Vezzali, L., Stathi, S., Giovannini, D., Capozza, D., Trifiletti, E., 2015. The greatest magic of Harry Potter: Reducing prejudice: Harry Potter and attitudes toward stigmatized groups. Journal of Applied Social Psychology 45, 105–121. https://doi.org/10.1111/jasp.12279

Vuletich, H.A., Payne, B.K., 2019. Stability and Change in Implicit Bias. Psychological Science 30, 854–862. https://doi.org/10.1177/0956797619844270

West, T.V., Magee, J.C., Gordon, S.H., Gullett, L., 2014. A Little Similarity Goes a Long Way: The Effects of Peripheral but Self-Revealing Similarities on Improving and Sustaining Interracial Relationships. J Pers Soc Psychol 107, 81–100. https://doi.org/10.1037/a0036556

Wilkins, C.L., Hirsch, A.A., Kaiser, C.R., Inkles, M.P., 2017. The threat of racial progress and the self-protective nature of perceiving anti-White bias. Group Processes & Intergroup Relations 20, 801–812. https://doi.org/10.1177/1368430216631030

Williamson, K.D., 2014. The Case against Reparations. National Review.

X, M., 1963. Racial Separation.

Young, I.M., 2000. Residential Segregation and Regional Democracy, in: Inclusion and Democracy. Oxford University Press, pp. 196–235. https://doi.org/10.1093/0198297556.001.0001

Zarya, V., 2018. The Share of Female CEOs in the Fortune 500 Dropped by 25% in 2018. Fortune.