

Forthcoming in *Ergo*
Alex Madva

A Plea for Anti-Anti-Individualism: How Oversimple Psychology Misleads Social Policy

Abstract

This essay responds to the criticism that contemporary efforts to redress discrimination and inequality are overly individualistic. Critics of individualism emphasize that these systemic social ills stem not from the prejudice, irrationality, or selfishness of individuals, but from underlying structural-institutional forces. They are skeptical, therefore, of attempts to change individuals' attitudes while leaving structural problems intact. I argue that the insistence on *prioritizing* structural over individual change is problematic and misleading. My view is not that we should instead prioritize individual change, but that individual changes are integral to the success of structural changes. These theorists urge a redirection of attention, claiming that we should think less about the individual and more about the social. What they should urge instead is that we think *differently* about the individual, and thereby think differently about the social.

I. INTRODUCTION

This essay responds to the criticism that contemporary efforts to redress discrimination and inequality are overly individualistic. Critics of individualism emphasize that these systemic social ills stem not from the prejudice, irrationality, or selfishness of individual actors, but from

underlying structural-institutional forces.¹ They are skeptical, therefore, of attempts to change individuals' attitudes while leaving structural problems intact. For example, Sally Haslanger writes: "Correcting the wrongs of racism, sexism, and the like is not best achieved by focusing on the 'bad attitudes' of individuals... structural injustices may persist even when attitudes change" (2015b: 2). Instead, Haslanger claims, "changing structures is often a precondition for changing patterns of thought and action and is certainly required for durable change" (1). Here I respond specifically to such arguments for *prioritizing* structural over individual reform in the struggle against prejudice, discrimination, and inequality.

Haslanger and others (whom I'll call structural prioritizers) have made invaluable contributions to our understanding of how group-based disadvantages can arise and persist in ways that do not depend on individuals harboring ill will towards members of other social groups. I agree that profound structural interventions are necessary for addressing these patterns of advantage and disadvantage, and I do not defend political or methodological individualism *per se* in any form. That said, I believe it is false and misleading to claim that we should prioritize structural over individual change (I'll call this misleading claim structural prioritism). I try to show that arguments for structural prioritism involve *non sequiturs* (§2) and that the policy predictions of structural prioritizers rely on oversimplified psychological models (§§3-4). Ultimately, the attempt to force a choice between structural and individual reform is confused. My view is thus not that we should *instead prioritize* individual change, but that individual changes will be integral to the success of structural changes.² Individual and structural changes

¹ See Alcoff (2010), Anderson (2010, 2012a,b), Ayala and Vasilyeva (2015), Banks and Ford (2009, 2011), Dixon et al. (2012), Haslanger (2015a,b), and Huebner (2016).

² See, for example, Antony (1995), Machery, Faucher, and Kelly (2009: §3), Christman's (2012) comment on Anderson (2012a), and Dotson (2012: 35-6).

are interdependent, although I conclude (§5) by describing what I take to be a kernel of truth underlying structural prioritizers' concerns.

My aim here is not to mark out a fundamental disagreement with structural prioritizers, but to urge friendly revisions to a specific and recurring cluster of positions that seem to occupy a central place in their criticisms of individualism. I hope to do justice, for example, to Haslanger's claim that "resistant agency" can "make a difference" (2015b: 8). Such endorsements of the difference-making power of resistant agency are, however, difficult to square with structural prioritism. I advocate holding onto the former and giving up the latter.

II. ARGUMENTS FOR PRIORITIZING STRUCTURAL CHANGE

The individual-level reform on which I focus is prejudice reduction.³ Many scientists and activists are exploring strategies for making individuals less biased against members of other social groups, and in particular, groups stigmatized on the basis of gender, race, ethnicity, sexuality, class, disability, religion, and so on. Structural prioritizers argue that this emphasis on prejudice reduction is misplaced:

The racial injustices that most trouble us are substantive—educational failure, large-scale incarceration, segregated and impoverished communities—and stem from a complex interplay of economic, historical, political and social influences. While historical bias has certainly played a role in producing these inequalities, it is fanciful to attribute their persistence to contemporary bias, unconscious or otherwise. The goal of racial justice efforts should be the alleviation of substantive inequalities, not the eradication of unconscious bias...

³ Criticisms of individualistic approaches arise in a variety of other spheres, e.g., positive psychology (e.g., Becker and Marecek 2008) and character development and education (e.g., Kohn 1997, Purpel 1997: 140, 146-52). My hope is that this paper represents a case study with potential ramifications for these other debates (see, e.g., note #11). See also Markus and Kitayama (2010) and Sarkissian (2010) for socially situated accounts of character which nevertheless emphasize the positive role of individual action and change, and Winton (2012) for insightful discussion of how to make character education less oriented toward individualistic virtues (as in, e.g., telling disadvantaged kids to act more "respectable") and more oriented toward structural change and social justice. Thanks to a reviewer at *Ergo* for urging me to consider the implications of my arguments for these parallel debates.

Pervasive racial inequalities persist... [but] the “problem” is those inequalities, not some supposedly biased mental state that has led to them... the now-dominant civil rights focus on mental state is misguided. (Banks and Ford 2011: 2, 13-14)

... the drive for prejudice reduction has for too long marginalized, if not obstructed, more pressing concerns about core distributive justice (e.g., justice based on the fair distribution of resources such as wealth, jobs, and health). (Dixon et al. 2012: 13)

The focus on individuals (and their attitudes) occludes the injustices that pervade the structural and cultural context and the ways that the context both constrains and enables our action. (Haslanger 2015b: 10)

According to structural prioritizers, focusing on individuals’ prejudiced attitudes⁴ occludes the primary causes of injustice, which are structural in nature.

Consider, for example, structural factors contributing to racial inequality. Elizabeth Anderson (2012a: 171) argues that “the structural ground” of persistent racial inequalities is not widespread racism, but ongoing *de facto* segregation in housing, education, employment, electoral districts, and so on. According to the “spatial mismatch” hypothesis, a major obstacle to employment for African Americans is the sheer distance between their homes and available jobs (Anderson 2010: §2.2). This difficulty is compounded when public transportation is inadequate, as it often is, and when jobs are advertised primarily by word-of-mouth, as many are. Such obstacles to black employment as spatial segregation and inadequate public transportation are structural in nature: they are features of the context or system in which individuals make decisions about where to seek jobs, which jobs to apply to, which job applicants to interview, and so on. These obstacles would persist even if individual employers tended to be unprejudiced

⁴ Writers like Banks, Ford, and Haslanger are responding in particular to the upsurge of interest in *implicit* (unreported) rather than *explicit* (self-reported) prejudice, while Dixon et al. primarily discuss explicit prejudice. I will not weigh in here on the various controversies about the best ways to measure, conceptualize, and predict prejudice and discrimination (but see, e.g., Madva 2016a,b; Brownstein and Madva 2012; Madva and Brownstein forthcoming). In what follows, I appeal to research on *both* implicit and explicit prejudice, and I believe we should draw the same general lessons for addressing both types of problematic attitude. Nevertheless, it is true that certain specific manipulations are more effective for changing implicit than explicit attitudes, and vice versa (see, e.g., Rydell et al. 2006; Moran and Bar-Anan 2013), so interventions must be designed to take these specificities into account. Thanks to a reviewer at *Ergo* for urging me to clarify these points. See also note #13 and §4, pp.23-5.

toward blacks, or even actively recruited and preferred black applicants. They would persist even if individual black jobseekers tended to be doubly motivated and qualified (“twice as good”) as white jobseekers. In a similar vein, as long as schools are funded by property taxes, segregation virtually guarantees unequal education.

Considerations like these lead Banks and Ford to suggest that it is fanciful to think that racism *per se* remains a major cause of persisting racial inequality. If racism vanished overnight, these structural factors, and therefore inequality, would persist. Other structural prioritizers, like Anderson and Haslanger, make the more moderate claim that pervasive (conscious or unconscious) racial prejudice and discrimination contribute significantly *less* to inequality than do structural factors. Notwithstanding these differences about whether racist attitudes and actions contribute “next to nothing” or just “not a whole lot” to ongoing inequality, structural prioritizers of all stripes draw similar practical conclusions: efforts to redress inequality should prioritize structural reforms—in this case, perhaps expanding public transportation, restructuring school funding, or renewing efforts toward racial integration—over individual reforms, such as trying to transform whites’ biases against blacks, or encouraging individual blacks to be more hardworking (or “respectable”). In fact, structural prioritizers tend to argue that racial prejudices are themselves primarily effects (“symptoms” or “mirror-like reflections”) of underlying inequalities (I say more about such claims in §§3-4). Anderson, for example, argues that segregation is a fundamental cause of prejudice: insofar as whites and blacks rarely interact, whites are less likely to appreciate the structural disadvantages blacks face, and more likely to (wrongly) attribute high rates of poverty and imprisonment in black communities to laziness, “gangsta culture,” or other stigmatizing generalizations. If racial prejudice vanished overnight, but segregation persisted, then segregation would eventually reignite racism. By contrast, if

segregation vanished overnight, and schools, neighborhoods, and voting districts were suddenly integrated, then racism would gradually decline. Thus, while Anderson (unlike Banks and Ford) acknowledges that racial biases make nontrivial contributions to persisting inequality, she thinks that integrationist structural reforms are themselves the best way to try to reduce those biases. Structural reforms should be prioritized, the reasoning goes, because they kill two birds with one stone, first by directly redressing the primary causes of injustice, and second by changing individuals' attitudes (a claim I return to in §3).

Anderson's discussion of segregation and integration exemplifies structural prioritizers' general analytic approach: to start with a backward-looking analysis of the primary, root causes of group-based disadvantages, and then to draw on this analysis to make forward-looking prescriptions about which strategies to prioritize to bring about change. I focus in this essay on the latter, forward-looking questions, but I will make occasional reference to backward-looking matters because structural prioritizers tie the two so closely together.⁵ "If racial segregation is the problem," Anderson argues, "it stands to reason that racial integration is the remedy" (2010: 112). Of course, it does not immediately follow from the fact that the primary cause of a given injustice is structural that we should prioritize a structural remedy. In general, there need not be any tidy, one-to-one, or in-kind mapping between problems and solutions. For example, it could be that climate change is the primary cause of the spread of a certain disease, but that the response to prioritize is to disseminate vaccines rather than to try to make the planet colder. It is an open-ended empirical question which specific interventions will be most conducive to which social changes. In light of this, structural prioritizers' recommendations are not deduced *solely* from their backward-looking analyses but also rely on empirical evidence, which they take to

⁵ Thanks to a reviewer at *Ergo* for emphasizing the importance of clearly and consistently distinguishing between the backward-looking and forward-looking questions. See also note #9.

demonstrate the weaknesses of individualistic reforms and the strengths of structural ones. Anderson (2010: Ch.6), for example, assembles a compelling body of evidence to suggest that integration may go a long way toward reducing racial injustice. In this section, I begin by assuming that structural prioritizers' backward-looking analyses as well as their empirical predictions are accurate. In §3, I reexamine some of these predictions, and, in §4, I suggest that the origins of structural prioritizers' mistaken predictions can be traced back to features of their backward-looking analyses. Specifically, I argue that structural prioritism relies on an oversimplified model of the mind, which leads to correspondingly misguided models of both the root causes of prejudice and the pathways to social change.

Suppose Anderson is right that integration should be the foremost structural remedy for combating racial inequality. Suppose Anderson is also right that integration is the best (most effective and just) way to reduce individuals' prejudices. Then perhaps integrationist structural reforms should be prioritized over independent attempts to reform individuals' prejudices, such as diversity training. That is, one particular reform (integration), which happens to be structural, should be prioritized over another (diversity training), which happens to be individualistic (at least in the sense that it aims primarily at changing individuals' attitudes and actions). But that is the most that Anderson's defense of integration, if it were correct, can show. It seems inaccurate and misleading to draw the more general conclusion that we should prioritize structural over individual reform full stop. Everything depends on *which* structural or individual reforms are at issue. First, there are numerous structural "reforms," such as overthrowing the government, that Anderson would obviously not advocate prioritizing over diversity training. Second, and more importantly, even if we grant that diversity training should be deprioritized, it does not follow that all forms of individual change should be deprioritized. In particular, *bringing about*

integrationist reforms will require, at a minimum, changes in the beliefs, motivations, or actions of those individuals poised to promote integration.⁶ Such 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 (a point I return to in §3). As a conceptual matter, these points may seem obvious, and I do not mean to suggest that structural prioritizers would necessarily disagree with them;⁷ as an empirical and practical matter, however, persuading and mobilizing individuals to implement and sustain structural reforms is no trivial task.⁸ The rhetorical emphasis of structural prioritism, that we must redirect attention from the individual to the social, tends to obscure the full scope and salience of these practical challenges.

So while in some cases it might make sense to say that we should prioritize certain structural over certain individual reforms (e.g., that we should prioritize integration over diversity training), I suspect that it will typically be more perspicuous to compare apples with apples and oranges with oranges by saying that we should prioritize certain structural reforms over others, and certain individual reforms over others. Regarding structural reforms, perhaps school integration should take precedence over revamping how schools are funded, and perhaps facilitating black movement into majority-white communities (for example, with housing vouchers) should take precedence over improving the housing and quality of life in majority-

⁶ Of course, the relevant structural changes *could* occur without any individual or collective efforts to bring them about, e.g., as an unintended benefit from some other change, but structural prioritizers are not pinning their hopes on such accidents. Also note that individual change can be necessary without being sufficient for structural change. To view individual change as sufficient would be to endorse a kind of (forward-looking) methodological individualism. But here my claim is that even if methodological individualism is false, that doesn’t make structural prioritism true.

⁷ But see Anderson (2010: Ch.9), which I discuss in note #19, and (2012b), which I discuss at the end of this section.

⁸ See, e.g., van Zomeren (2013) for a helpful overview of individual-psychological variables relevant to initiating structural change.

black communities.⁹ Regarding individual reforms, perhaps *motivating individuals to support integration* should take precedence over reducing individuals' prejudices. Note, however, that for this latter contrast between types of individual reform to make sense, we would have to assume that the most effective strategies for mobilizing support for racial integration would not include a significant role for prejudice reduction, which seems odd. Intuitively, we might expect that reducing pervasive racial prejudice would be a pretty good way to drum up enthusiasm for integration—in which case changing the bad attitudes of individuals might be integral to bringing about the structural reforms that are integral to promoting justice. That is, even if we grant that the struggle for racial justice should above all prioritize integrationist structural reform, it still does not follow (a) that we should deprioritize individual change in general, *nor even* (b) that we should deprioritize individual prejudice reduction in particular. Our initial inference, that prioritizing integration entails prioritizing it *over* independent attempts to reduce prejudice, now seems open to question.

However, John Dixon and colleagues (2012) further argue that reducing prejudice, or, as they put it, “getting individuals to like each other more,” is not just ineffective but even counterproductive for promoting justice. They target the social contact hypothesis, which predicts that intergroup interaction reduces prejudice, especially when members of different groups frequently cooperate toward a common goal on equal terms. Meta-analyses generally confirm this prediction (Pettigrew and Tropp 2006). Recently, however, a spate of studies suggests that extensive contact between high-status and low-status groups also has some

⁹ See Shelby (2014) for trenchant discussion of Anderson's arguments for the primacy of segregation *qua* cause of inequality and the primacy of integration *qua* redress. Shelby defends an “egalitarian pluralist” approach to bringing about racial justice that rejects the prioritization of integration over other reforms, such as the improvement of black neighborhoods. Shelby claims that black residential self-segregation is often permissible while white self-segregation is typically not.

unintended, counterproductive effects. In particular, it leads members of low-status groups (including blacks in South Africa, Arabs in Israel, Muslims in India, and black college students in the US) to perceive the status quo to be fair, to be less supportive of structural reform, and to (often mistakenly) expect fair treatment from members of high-status groups. “When the disadvantaged come to like the advantaged, when they assume they are trustworthy and good human beings, when their personal experiences suggest that the collective discrimination might not be so bad after all, then they become more likely to abandon the project of collective action to change inequitable societies” (Dixon et al. 2012: 11). In short, social contact leads them to like the advantaged group more, but also saps their motivation to fight for social change.

Dixon and colleagues attribute the failure to anticipate and reckon with these counterproductive effects to an “individualist focus” in prejudice reduction research, which, they claim, is organized around the prediction that:

By changing individuals’ prejudices, we also change how they relate to other people in their lives, and in turn this effect is believed to ripple outwards to shape wider patterns of intergroup conflict and discrimination... (2012: 7-8)

They contrast this putatively individualistic model to a:

collective action model... [for] achieving social justice. Its guiding assumption is that social change is predicated upon mass mobilization, a process that typically brings representatives of historically disadvantaged groups (who stand to benefit from change) into conflict with representatives of historically advantaged groups (who stand to lose out from change). (8)

Similarly, Haslanger (2015b: 12) writes, “Social change requires contestation, organization, and activism.” She implores us to ask not, “what should I do?” but “what should we do?” (11).

These theorists make a number of important contributions to our understanding of discrimination, inequality, and social change. Again, however, the broader conclusion about the relative priority of individualistic versus anti-individualistic reforms is a *non sequitur*. First, it is

simply inaccurate for Dixon and colleagues to portray proponents of intergroup contact as individualist. Dixon and colleagues suggest that the model underlying the contact hypothesis is: change sufficiently many individual hearts and minds, who will then act in less discriminatory ways, which will then lead to large-scale structural changes.¹⁰ This is not the model. Positive intergroup contact is not predicted to spring up *ex nihilo*. Contact theorists argue that numerous structural interventions, such as integrating neighborhoods, schools, and voting districts, are necessary to make positive intergroup contact possible. Dixon and colleagues inexplicably elide these structural interventions in their portrayals of the contact hypothesis. In fact, contact theorists—such as Anderson—tend to be structural prioritizers. Dixon and Anderson merely disagree over which structural reforms to prioritize.

Inextricably intertwined with questions about which structural changes to prioritize are questions about which individual changes to prioritize. One question that the data assembled by Dixon and colleagues requires us to ask is, roughly, which of the following two changes will better promote social justice: either lots of individuals become less prejudiced, or lots of individuals (become motivated to) fight for social change. If Dixon and colleagues' assessment of the data is accurate, we can't do both, because prejudice reduction brings political malaise in tow, and political activism requires (a certain kind of) intergroup discord. We must, then, reexamine which factors strengthen individuals' motivation to fight for change. Presumably this

¹⁰ See, for example, how Wright and Baray (2012: 236) contrast what they call the “causal flow” through three “levels of analysis” in individualist versus anti-individualist models, where the micro-level of analysis refers to “intrapersonal phenomena (cognitions, emotions etc. within individual people),” meso-level refers to “interpersonal phenomena (the interactions between individual people), and macro-level refers to “broader social institutions and the structural relations between groups in societies”: “Prejudice reduction seeks to alter micro-level phenomena to spark reductions in interpersonal acts of discrimination (meso-level change), and as these more positive interpersonal behaviours proliferate, they should reduce macro-level structural and status inequalities. The collective action model describes actors' efforts to change the macro-level intergroup relations as the starting point, and to the degree that it is concerned with interpersonal (meso-level) and intraindividual (micro-level) phenomena at all, this view would propose that greater equality at the macro-level might then alter the dynamics of interpersonal cross-group interactions and individual psychological processes.”

includes institutional, structural, and social factors, such as a national voting holiday, accessible public transportation to voting stations, and peer pressure to march and vote—and also psychological factors, like the desire to effect change and the belief that marching and voting will be instrumental to doing so.

Indeed, when structural prioritizers redirect our attention to phenomena more conducive to social change, we find, *inter alia*, claims about specific individual-psychological dispositions to target. Dixon and colleagues claim that we should encourage individuals “to recognize injustice and status disparities and thus strive to change the status quo” (9).¹¹ In light of such claims, some central empirical questions for Dixon and colleagues would presumably be how best to get individuals to appreciate that certain group-based disadvantages are unjust, and to motivate them to act in response. One example of (individualistic?) psychological research relevant to these aims might be Johnson and Fujita’s (2012) finding that making the malleability of a social system salient increases people’s interest in the negative aspects of that system and also increases their motivation to change the system.¹² That is, if you make individuals aware that a certain social system *can* be changed, then they are more likely to focus on *what* to change, and become more motivated to change it. Perhaps changing individuals’ attitudes about the malleability of social structures is more important than changing individuals’ attitudes about other social groups.

¹¹ See Becker and Maracek (2008: 1773-4) and Winton (2012: 43-4, 55-8) for similar points in the context of positive psychology and character education, respectively. Dixon and colleagues also claim that we should promote “common identity” and political solidarity between members of disadvantaged groups rather than between members of disadvantaged and advantaged groups. Far from requiring radical revisions to prevailing theoretical models or research programs, such aims fit rather seamlessly into the traditional prejudice reduction paradigm. Creating political solidarity between disadvantaged groups likely requires something an awful lot like cooperative intergroup interaction, i.e., exactly the kind of social contact that has been predominantly studied between advantaged and disadvantaged groups. But this is speculative; part of the problem is that researchers may have been focused on the *wrong individuals*, i.e., the advantaged instead of the disadvantaged. More generally, we should expect that the sorts of changes individuals should prioritize will vary greatly depending on the specifics of their social locations.

¹² See also Stewart et al. (2010).

In a companion paper (ms.), I describe a variety of strategies that individuals can pursue to transform their cognitive, affective, and motivational dispositions. My focus there is on interventions that reduce prejudice, stereotyping, and discrimination, but nothing suggests that these techniques are content-specific and only applicable to particular psychological dispositions. Many effective prejudice reduction techniques are explicitly modeled on clinical interventions, for example, for ameliorating phobias and addictions. It seems that, by simply varying the stimuli, the very same experimental techniques can reduce relapse among recovering addicts (Wiers et al. 2011; Eberl et al. 2013), reduce the influence of stereotypes on hiring decisions (Kawakami et al. 2007a), improve interracial social interactions (Kawakami et al. 2007b), eliminate stereotype threat (by increasing motivation and performance on math tests; Forbes and Schmader 2010), increase self-esteem, and change preferences for *Haribo* gummy bears versus *Milka* chocolate bars (Ebert et al. 2009).¹³ Perhaps individuals can use such techniques to increase their motivation to vote and march, or to cultivate an egalitarian ethos. Of course, increased avidity for political participation and enthusiasm for egalitarianism will not suffice to end social inequality, but structural prioritizers give us no reason to think that these psychological changes are somehow irrelevant to that end.

¹³ A reviewer for *Ergo* notes that different sorts of intervention might affect individuals differently depending on developmental factors. For example, perhaps “individualist interventions work best for children and young people, but once a person’s core character has more or less been set in stone, only structural interventions will carry significant force.” I strongly agree that interventions will most likely have to be tailored to the specific developmental levels of the individuals involved. That said, one of the most striking upshots of the research cited here and in §3 is the extent to which it upends a certain intuitive view of prejudices, namely, as so deeply ingrained during childhood that they become difficult to change in adulthood (for further discussion, see Madva ms.). This intuitive view is mistaken about both implicit and explicit attitudes: the research I cite in §3 regarding exposure to counterstereotypes exemplifies how both types of attitude are more malleable and sensitive to situational contingencies than previously realized. Nevertheless, to the extent that young minds are more flexible than older ones, I would likely advocate implementing *both* individualized interventions (like the prejudice reduction techniques cited here) and structural interventions (like integration) early and often in children’s education.

The point clearly generalizes. It is difficult to see how general claims about prioritizing structural over individual reform could be true, unless structural prioritizers are envisioning some way of effecting structural reform in complete independence of individual participation. But structural prioritizers are clearly trying to *reorient* our modes of political participation, not bypass participation altogether.

That said, Anderson at times nearly claims that racial justice *can* be promoted without individuals fighting for it. In response to John Christman (2012), who raises concerns similar to mine, Anderson effectively doubles down on structural prioritism. She states, for example, that “the means by which integration works to promote justice operate largely behind people’s backs” (2012b: 1). The prediction, roughly, is that cooperative interracial interactions will reduce bias and enrich blacks’ social capital, economic opportunities, etc., even if affected individuals are unaware that these changes are occurring. Even here, Anderson acknowledges that, in keeping with the contact hypothesis, integration promotes justice when there is “hierarchical enforcement of intergroup cooperation by officers, coaches, and bosses” (1). That is, (individual) leaders of institutions must put the integrationist structural changes in place, vocally endorse them, and enforce them. Subordinate individuals must also defer to these integrationist authority figures: “deference to authority can substitute for direct endorsement of integration.” So deference to authority would be one individual-psychological disposition amenable, at least in this context, to advancing social justice. Of course, deference to authority is not always a good thing (for example, when authority figures are Nazis, or segregationists, or experimenters instructing participants to deliver electric shocks to other people), and it is not always easy to come by (for example, when federal interventions are perceived as tyrannically overriding the rights of states and local communities to self-govern). But if Anderson were right, then some salient empirical

questions would be which interventions lead individuals in positions of authority to endorse integration, which interventions lead individuals to defer to the authority of their integrationist leaders, etc. Again, the takeaway would not be to deprioritize individual change in general, but to deprioritize some individual changes and prioritize others.

III. STRUCTURAL CHANGE AS PRECONDITION FOR INDIVIDUAL CHANGE

I assumed in §2 that structural prioritizers' empirical predictions were accurate, but I will now examine these predictions more closely. Structural prioritizers tend to argue not just that structural change is more important than individual change, but that structural change will itself be the best way to effect lasting individual change. As Haslanger says, "changing structures is often a precondition for changing patterns of thought and action and is certainly required for durable change" (2015b: 1). Structural reform is thought to kill two birds with one stone, first by directly redressing the primary causes of injustice, and second by changing individuals' attitudes. A paradigmatic example of this sort of argument appears in recent defenses of affirmative action. Structural prioritizers argue that affirmative action directly counters ongoing discrimination and patterns of (dis)advantage, but that it also reduces prejudice. They claim that promoting members of underrepresented groups to positions of prominence will produce *debiasing agents*, counterstereotypical exemplars who will reduce the prejudices of their peers.¹⁴ Anderson writes:

Through demonstrably successful functioning in their roles, the targets of affirmative action help break down racial stereotypes that underlie stigmatization and discrimination. This effect is not simply a matter of bringing counterstereotypical individuals to the attention of other participants in institutions practicing affirmative action. On a larger scale, affirmative action aims to break the public association of blacks with poverty and

¹⁴ See, for example, Anderson (2010, 2012a,b), Dasgupta (2013), Huebner (2016), Jolls and Sunstein (2006), and Kang and Banaji (2006).

associated dysfunctional behaviors by moving blacks to secure middle-class positions, reproduced across generations. (2010: 150)

Matters are not so simple. Claims like these fail to appreciate the myriad complexities and contingencies involved in how individuals interpret and react to structural interventions.

What sorts of empirical evidence are thought to license these claims about affirmative action's debiasing power? Structural prioritizers emphasize findings that counterstereotypical teachers debias their students in long-lasting and wide-ranging ways. For example, undergraduate women are more likely to like and pursue math and science—and, more generally, to think of themselves as assertive leaders with significant career goals—if they have a few women math and science professors (Dasgupta 2013; Dasgupta and Asgari 2004; and Stout et al. 2011). However, the centerpiece of the structural prioritist defense of affirmative action as a two-bird stone is that it will change the biases of members of *other* social groups (Kang and Banaji 2006: 1109-10), but research increasingly suggests that having women professors has *no effect whatsoever* on undergraduate men's implicit or explicit stereotypes about math ability and gender. This particular debiasing effect applies only to ingroup members (women) rather than outgroup members (men).¹⁵ Moreover, even this ingroup effect depends on several contingent psychological factors, such as the extent to which individuals perceive themselves to be similar to the counterstereotypical exemplar. For example, the effect increases when participants believe that the exemplar graduated from their own university. But if women believe that the exemplar is an exceptional “superstar” genius, then the effect *reverses*: they report fewer career aspirations, think of themselves as less assertive, and lose interest in math and science (Asgari et

¹⁵ For the record, I think such ingroup role-modeling is very valuable. Anecdotally, I hear that defenders of affirmative action have distanced themselves from the ingroup role-modeling defense of affirmative action in large part because it has been shot down by the courts.

al. 2010, 2012). It is simply false that mere exposure to counterstereotypes, just as such, ensures debiasing effects, even for ingroup members. The effects are hostage to a variety of highly contingent factors, many of which, including perceptions of similarity, are *psychological* in nature. The effects depend on how specific individuals interpret and react to their social environments. Individual change does not come “for free” when necessary structural reforms are put in place.

Anderson claims that the “demonstrably successful functioning” of minorities in prominent roles will “break down” stereotypes. But *successful performance* is often not an unmistakable social fact that gets transmitted directly and unfiltered into others’ minds. The very stereotypes in question could prevent others from recognizing an individual’s successful performance as successful.¹⁶ Moreover, if coworkers *believe* that someone has been promoted ahead of them simply to satisfy a quota (or, for that matter, to advance social justice, rather than because they are the best candidate for the job), they may resent what they (perhaps wrongly) perceive to be undue benefits. The mere presence of diversity-promoting structures (which may or may not be effective) can create an “illusory sense of fairness” among privileged individuals, who in turn become *more discriminatory* (Kaiser et al. 2012). The perception that certain individuals have benefited from affirmative action or other structural interventions can lead supervisors and co-workers to “compensate” for this perceived benefit, subsequently treating them worse and under-evaluating their performance.

Even if Anderson were right that the successful functioning of these individuals helps to, as she says, “break the public association of blacks with poverty,” the overall consequences of such a revision in public perception are likely complex. Consider some of Michelle Alexander’s

¹⁶ See, for example, Eagly and Karau (2002) and Valian (1998). See Kukla (2014) for insightful discussion of similar cases.

concerns about affirmative action and counterstereotypical exemplars:

Highly visible examples of black success are critical to the maintenance of a racial caste system in the era of colorblindness. Black success stories lend credence to the notion that anyone, no matter how poor or how black you may be, can make it to the top, if only you try hard enough. These stories “prove” that race is no longer relevant. Whereas black success stories undermined the logic of Jim Crow, they actually reinforce the system of mass incarceration. Mass incarceration depends for its legitimacy on the widespread belief that all those who appear trapped at the bottom actually chose their fate.

Viewed from this perspective, affirmative action no longer appears entirely progressive. So long as some readily identifiable African Americans are doing well, the system is largely immunized from racial critique. (2010/2012: 248)

If Alexander is right, then policies focused on advancing the status of elite African-Americans might not be “entirely progressive,” and might even play an overall counterproductive role, serving to reinforce racial inequality rather than undermine it, by concealing the disadvantages faced by the worst off. The once prevalent claim that the election of a black President signaled that we had entered a post-racial era is a case in point. If Alexander is right, then structural reforms like affirmative action indeed make strong impressions on individual minds, but the net effect, in this case, would be to conceal more injustice than it combats.¹⁷

As further evidence for backlash effects of structural reforms on individual attitudes, take the recent case of MIT.¹⁸ In 1999, MIT found that it had been systematically discriminating against women, e.g., by allocating them less lab space than men. The administration took responsibility for redressing gender discrimination and a 2011 report found that “stunning progress” was made (10). In twelve years, the number of women faculty doubled, more women occupied leadership positions, and almost no gender differences in average salary, lab space, and

¹⁷ Insofar as blacks are more likely to be poor, “the public association of blacks with poverty” is not in itself a bad thing. This association reflects a real-world injustice that we need to remain aware of in order to combat. Hence, the goal of prejudice reduction should not be to *eliminate* this association from our minds altogether. The ethical and political problems of this association depend on when and how it comes to our minds, and in how we are inclined to explain it. Problems arise when the association springs to mind even when it is irrelevant, or when we are inclined to explain black poverty in terms of the biological or cultural traits of people of African descent, rather than in terms of ongoing discrimination and structural disadvantage. See Madva (2016a) for further discussion.

¹⁸ “A Report on the Status of Women Faculty in the Schools of Science and Engineering at MIT,” (2011).

teaching loads remained. Nevertheless, many women report that they still battle gendered expectations for behavior (to be “neither too aggressive, nor too soft” (16)), that childcare is still perceived as a “women’s issue,” and that there is now a widespread perception that they have unfair advantages in hiring and promotion. Despite important structural reforms, individuals’ gender biases persist and in some respects, like perceptions of unfairness, got worse. Again, psychological change does not come for free along with structural change.

There is, moreover, little reason to count on those perceptions of unfairness to somehow gradually fade with time.¹⁹ They may instead grow, and eventually motivate efforts to roll back the structural reforms. Zooming out from the case of MIT, there is ample historical evidence of high-status groups perceiving structural reforms as unfair, remaining and perhaps becoming increasingly hostile to those reforms over time, and ultimately working to undo them. This pattern is clearly visible in much of the ongoing opposition to affirmative action, just as it was and remains visible in white resistance to active efforts toward desegregation, and in the “Redemption” of white power in the south following the post-Civil-War Reconstruction.

We cannot simply count on structural interventions to be two-bird stones, simultaneously redressing basic conditions of injustice and transforming individual hearts and minds. To the contrary, implementing structural reforms without sufficient understanding of or attention to the

¹⁹ See, for example, Anderson’s (2010: 121-2, 126, 156, 182-3) arguments that backlash to structural interventions is to be expected but will dissipate over time. Anderson also argues that white resistance to integration is unjustified and based on unfair stigmatizing ideas (155-6), and therefore “its normative force should be drastically discounted” (170). I agree that the resistance of the privileged to structural reform is typically unjust but it is problematic for Anderson to invoke this point in defense of structural prioritism. To begin with, Anderson’s theory *predicts* that segregation *causes* whites to resist integration. It is somewhat awkward to give a comprehensive causal explanation of why people tend to endorse a certain problematic view and then to insist that they should not “be let off the moral hook” on the grounds that it would not be “unreasonably difficult” for them to reject that view (190), especially since Anderson offers a considerably more nuanced analysis of the moral responsibility disadvantaged blacks bear for the problematic dispositions that segregation causes them to have (§4.3). In any case, debating the normative status of these attitudes is a red herring, in this context. Whether or not backlash is justified, it is entirely predictable, and therefore poses an obvious problem for structural prioritism. Tactically speaking, it just doesn’t make sense to say a certain intervention should be prioritized (and to criticize other proposed interventions for being “unrealistic”) if we predict that widespread resistance to the intervention will make it “politically unfeasible” (189).

motivations, interpretations, and biases of the individuals involved can easily backfire, begetting heightened prejudice and discrimination. As I see it, the upshot is *not* that interventions like affirmative action are a bad idea. Affirmative action serves vital social goods. Rather, we must figure out which individual attitudes facilitate the acceptance of affirmative action, and then figure out how to promote those individual attitudes.²⁰ Discrimination and inequality are *two-stone birds* that must be jointly tackled by individual and structural reforms. Earlier (§2), I suggested that it would be more appropriate to urge prioritizing *among* candidates for structural reform and *among* candidates for individual reform, rather than *between* structural and individual reforms. More appropriate still would be to urge prioritizing among symbiotic bundles of individual-structural reforms, wherein changes in structures and in hearts and minds work in tandem. Which interventions are most conducive to encouraging individuals to understand and mobilize against structural injustice? Which structural reforms are least likely to beget counterproductive backlash and most conducive to producing durable and desirable changes in individual hearts and minds?

To be fair to structural prioritizers, many studies do show that exposure to counterstereotypes can reduce individuals' biases. I say more about these studies elsewhere (ms.), but many of the most promising studies require *individuals to invest some effort* in the interventions that reduce their prejudices, for example by actively imagining, attending to, and affirming counterstereotypes. In many cases, participants are not merely passively exposed to counterstereotypes in their environment, but are active in endorsing them *qua* counterstereotypical. These studies are, in other words, paradigm cases of individual-level

²⁰ See Mizell (2012) for discussion of factors that influence individuals' attitudes toward affirmative action. For a striking example of how to persuade white Americans to buy into integrated schools, listen to Hannah-Jones and Joffe-Walt (2015, August 7), "The Problem We All Live With, Part Two," *This American Life*.

remedies, where individuals deploy conscious strategies and interact with their environments to become agents in their own debiasing. These studies make clear the active role that individuals can take in interpreting and reacting to their environment, and the beneficial consequences such activity can bring.²¹ This active role is elided in structural prioritizers' models and policies for redressing injustice.

IV. PSYCHOLOGICAL BIASES AS MIRRORS OF SOCIAL REALITY

Anti-individualists' confidence about the priority of structural over individual reform, and about the power of structural reforms to transform individuals' hearts and minds, reflects, I think, an unduly passive view of hearts and minds. Although the primary focus of this paper is the forward-looking question about which strategies to prioritize, the mistake in structural prioritizers' view of the mind is most clearly seen by tracing it back to its origins in their backward-looking analyses, and specifically in their model of the root causes of prejudice. For example, social psychologist Nilanjana Dasgupta (2013: 240-1) claims that prejudices and stereotypes are:

mirror-like reflections of local environments and communities within which individuals are immersed... implicit preference for some groups and bias against others are learned associations acquired by passive immersion in an unequal society where people are segregated into disparate roles, jobs, and geographies based on group membership. In everyday life and in daily media, people observe that some types of individuals typically occupy highly valued roles, while others typically occupy devalued roles, and this distinction is often based on group membership... Through repetition, these observations get passively recorded in the mind and become the basis of implicit attitudes and beliefs.

Dasgupta here offers a representative description of how unjust social structures produce biased

²¹ See Sarkissian (2010) for insightful discussion of similar pitfalls in the situationist critique of virtue ethics.

minds, according to structural prioritizers.²² Call this view MIRROR. Specifically, to say that our biases are “mirror-like reflections” of the social world is to say that they are acquired and sustained simply by virtue of the fact that we grow up and remain immersed in a society structured by visible disparities between social groups. We are passively, involuntarily, and perhaps unconsciously socialized into acquiring these undesirable biases by being bombarded with stereotypes in mass media and by observing that people from specific social groups are more likely to occupy specific roles, have specific jobs, live in specific areas, and so on. There are roughly two primary causes of bias on this view: first, knowledge of *genuine* disparities between social groups, and second, repeated exposure to *distorted* representations of social groups. In both cases, the mental absorption of bias is fundamentally passive and unavoidable: “*mere knowledge* of stereotypes, even if they are consciously repudiated, is sufficient to bias behavior toward stereotyped groups” (Glaser 2015: 61). A mirror merely reflects the light that hits it. It does not produce any light of its own.

MIRROR plays a key role in guiding empirical predictions about how structural changes will lead to individual changes. Dasgupta continues:

If this is how implicit attitudes develop then such biases should shift when people are immersed in different types of situations where they encounter admired and counterstereotypic individuals who do not fit their prescribed role in society... Changes in these environments and communities... produce changes in implicit attitudes and beliefs... (241, 240)

Since social biases are modeled as *effects* rather than *causes* of social environments, changing those environments should lead to changes in bias—and not the other way around. As long as

²² See, for example, Huebner (2016) for an approving reference to Dasgupta. In some very general sense, it is obvious that social biases depend on social environments. We couldn't have biases against social groups that we never knew existed. Of course, in another general sense, the relevant structures and environments obviously depend on individual minds. Money and power would not exist if there were no minds who believed in them. My concerns are with structural prioritizers' specific models of *how* social phenomena influence psychological phenomena.

unjust structures are in place, it will be more or less futile for individuals to devote significant efforts toward reducing their own prejudices. Bryce Huebner (2016: 71) writes:

So as we watch or read the news, watch films, rely on tacit assumptions about what is likely to happen in particular neighborhoods, or draw elicited inferences on the basis of the way in which a person is dressed, we cause ourselves to backslide into our implicit biases. No matter how calm, vigilant, and attentive to our biases we try to be, I maintain that we will be unable to moderate or suppress all of our problematic implicit biases until we eliminate the conditions under which they arise.

Such predictions are a point of emphasis in defenses of structural prioritism. We can roughly capture their empirical predictions in two claims, which reflect a kind of causal asymmetry: first, it is not the case that changing individuals' biases will cause significant and enduring changes in social structures, and second, it is the case that changing social structures will cause significant and enduring changes in individuals' biases. MIRROR underwrites these predictions. If you think that biases are mirror-like reflections of social environments, then it is pointless to try to change biases without changing the environment, and, if you do change the environment, then changes in biases will follow suit.

The funny thing about MIRROR is that it is completely wrong and everybody knows it. We are not empty heads that just get filled with the preponderance of information we encounter. It is old news that we don't work like that. There are myriad ways in which our beliefs and expectations are not closely calibrated to the actual regularities we encounter. We suffer from a profound confirmation bias, being more likely to seek out and attend to evidence that reinforces what we already believe than to consider contravening evidence. Moreover, our beliefs often persevere in the face of the contravening evidence that we *do* happen to consider. For example, Lord and colleagues (1979) found that individuals' attitudes about the death penalty only became

stronger and more polarized when they confronted powerful evidence for the opposing view.²³ Similarly, Handley and colleagues (2015) found that men were more critical than women of empirical evidence suggesting a bias against hiring women in the sciences (whereas women were more critical of evidence suggesting the absence of such a bias). Optimism that our biases will simply melt away after exposure to sufficiently many counterstereotypes seriously underappreciates the human capacity to interpret the facts in ways that prop up what we already believe. This capacity, it turns out, is just as true of our implicit attitudes and beliefs as it is of our consciously articulated hypotheses (Valian 1998). (It is not even clear the extent to which these cognitive dispositions are irrational, given that empirical evidence paradigmatically underdetermines theory.)

Of course, I do not mean to imply that being bombarded with stereotypes and being immersed in a society marked by visible disparities between social groups is *irrelevant* to the acquisition and maintenance of bias, nor to imply that being bombarded with counterstereotypes is irrelevant to the dissolution of bias. The point is that it is not *simply* by virtue of repeated exposures that biases form and change. There are an awful lot of contingencies involved, many of which have to do with individual psychology. In particular, what we notice and how we interpret our social environments is profoundly shaped by our implicit and explicit goals (Kunda and Spencer 2003; Moskowitz 2010; Uhlmann, Brescoll, and Machery 2010; Madva 2016a). Goals that work in favor of stereotyping include the desire to protect one's self-esteem (for example, by putting down another group) and to see the world as a fundamentally just place where people deserve their lot. So, for example, when we encounter a distorted and stereotypical representation of an outgroup, part of what might lead us to think, "There's a grain of truth in

²³ See Kenyon (2014) for discussion of polarization biases and arguments for prioritizing social over individual reforms to address these biases.

that,” is that doing so makes us feel better about ourselves, and better about the overall fairness of the world, not just that we have been exposed to many examples of this stereotype in the past. By the same token, other goals, such as to be egalitarian or creative, can also be recruited to *resist* stereotyping.

MIRROR is a radically oversimplified and misleading gloss on the psychology of prejudice (and the psychology of everything else), and is belied even by the data of psychologists who make such claims. It is especially puzzling to see Dasgupta invoke MIRROR because, in the very same paper, she reviews a wide range of her own pioneering research that speaks against it, including several of the studies I mentioned in §3 demonstrating some of the contingencies involved in how counterstereotypical professors influence their students (the debiasing effect sometimes only works on ingroup and not outgroup members, it depends on how similar the individual feels toward the professor, etc.). Dasgupta’s (2013, 258) *actual* model of how individuals respond to counterstereotypical role models is far more complex than her gloss of biases as “mirror-like reflections” would suggest.

MIRROR reflects a pervasive, longstanding, commonsensical view of prejudice, which has—as beliefs so often do—persevered despite being thoroughly debunked, and now continues to inform how social scientists, activists, and philosophers think about strategies for redressing prejudice, discrimination, and inequality. Few theorists actually come out and endorse MIRROR wholeheartedly. The view comes in more and less explicit forms, and is more likely to rear its head in certain contexts than others. Specifically, structural prioritizers including Anderson (2010, 2012a), Haslanger (2015b), and Huebner (2016) tend to be much better at moving past MIRROR and appreciating the complex interplay of structural and psychological factors when they are explaining the perpetuation of injustice than when they are envisaging the pivot toward

justice. They seem to be much better at characterizing vicious cycles than virtuous cycles. Whence this asymmetry? I think they are rightly identifying flaws in leading proposals for redressing social injustice, but that these flaws typically have little to do with individualism.

**V. THE KERNEL OF TRUTH BEHIND STRUCTURAL PRIORITIZERS' CONCERNS:
THE LAW OF UNINTENDED CONSEQUENCES**

When it comes to combating these social ills, we are learning that matters are more psychologically complex than we realized. We are also learning that matters are more socially complex than we realized. One important insight behind structural prioritism is that individual change alone will be insufficient to address these multifaceted problems; structural change is clearly necessary as well. I think a further kernel of truth behind the claim that we should prioritize structural over individual change is, roughly, the “law” of unintended consequences: well-intended interventions into complex systems often have unforeseen effects.

One way of bringing this out is to return to debates among structural prioritizers regarding which specific interventions to prioritize (§2). Anderson argues that integration is the structural intervention *par excellence* both for redressing inequality and for reducing prejudice. Others, such as Dixon and colleagues (2012), criticize such views for being too individualist. This criticism is misguided: Dixon and colleagues are inexplicably discounting all the structural interventions that enable positive intergroup contact. (I argued, however, that structural prioritizers across the board are inexplicably discounting all the individual-level remedies that enable the structural remedies that enable the individual remedies, and so on. Anti-anti-

individualism is about keeping the complex interface between individuals and structures in view not just in the context of theorizing injustice but in theorizing paths to justice.)

While the criticism of the contact hypothesis as individualist is off-base, the empirical evidence that sparked this criticism warrants serious attention. This is the kind of evidence I discussed in §2 of counterproductive effects of prejudice reduction. The real headline is that the contact hypothesis, which ranks among the most extensively studied and widely praised theories in all of the social sciences, may have a whole bunch of unintended negative consequences. The widespread failure to detect these consequences sooner does not reflect an excessive focus on individual minds, but it does reflect an excessive focus on the wrong parts of individual minds, and specifically, a failure to consider the downstream effects of an intervention that effectively reduces prejudice. It constitutes a failure to be on the lookout for unintended consequences.

Sometimes this failure is due to deep-rooted methodological biases in our scientific practices, such as testing a medicine only on men and not on women. Other times, bad outcomes happen to good theories, and there is neither any ideology to blame for the error nor any need to go back to the theoretical drawing board. The basic lesson might be a familiar one, that human minds, like social systems, are terribly complex. When we try to influence complex systems, there is a methodological imperative to search for unintended side effects. Make sure you're curing the disease without killing the patient. Of course, since there will always be "unknown unknowns," there is no way to guarantee that an intervention will be free of unintended consequences. It is a scare-quotes "law" because it is ubiquitous. There is, then, a second methodological imperative, to adapt flexibly when those unintended consequences become evident, as they eventually do. In some cases, the unanticipated side effects may be so disastrous that the entire intervention should be abandoned. In other cases, the side effects may be avoided

or minimized with a less drastic revision or complement to the intervention. When treating the disease induces nausea, take anti-nausea medication.

One of the most common criticisms made by structural prioritizers of all stripes is that their opponents' proposals for redressing discrimination and inequality will obscure or marginalize the root structural problems. That is, putting such-and-such policy in place will have the undesirable unintended consequence of distracting us from more pressing issues. For example, Anderson opposes cash reparations to redress past racial injustice on two grounds:

First, allocating lump-sum reparations to blacks is like serving water to the thirsty in a sieve. Unless the continuing causes of race-based disadvantage are dismantled, such reparations will only offer temporary relief. Second, reparations' focus on compensating for injustices in the past distracts attention from current injustices and is liable to encourage whites to feel that, once paid, they have done everything needed to end racial injustice, and to place all responsibility for continuing racial inequality on blacks alone. (2010: 229, n.19)

I withhold judgment here about the relative efficacy of cash reparations versus Anderson's preferred strategy of integration, but her argument—that reparations would not just fail to ameliorate underlying structural injustices but would even distract us from those injustices, and lead whites to believe that their moral work was done because the debt of racial injustice had been paid—is unfair and rhetorically risky. Perhaps the most ubiquitous side effect of interventions, individual or structural, effective or ineffective, for combating discrimination and inequality (and probably for combating anything else) is that their very implementation will tend to reduce people's sense of urgency in solving the problem (Mann and Kawakami 2012).

Whenever we think we are making progress, we become less motivated to keep fighting. It is unfair, then, to hold this unintended consequence against any view in particular. It is also unfair to criticize an intervention on the grounds that implementing it *in isolation* from other interventions would fail to uproot the underlying causes of injustice. I see no reason why

reparations and integration should be construed as competitors. Why can't wealth redistribution figure alongside integration as two important parts of a comprehensive strategy for promoting justice?²⁴ The fact that reparations would not solve all forms of inequality all by itself is no argument against its figuring as an important component of a broader strategy. On this point, one might be tempted to respond that it would be unfeasible to both give reparations and promote integration, e.g., due to limited resources. I have two responses here. First, I am skeptical about appeals to limited resources in this context, because the resources available are not drawn from a fixed cache but depend on protean political, economic, and psychological factors. Second, the appeal to unfeasibility and limited resources is plainly unavailable to structural prioritizers, who argue that many different and intensive structural reforms are needed to effectively combat discrimination and inequality. Integration, for example, is not a single intervention but a composite of different interventions, each of which would make separate demands on our limited time, attention, and material resources. Anderson acknowledges that pursuing only one integrationist reform in isolation, such as busing across school districts, might be counterproductive, such as by inducing white flight (2010: 190).

Anderson's argument is also rhetorically risky. After all, Dixon and colleagues argue that similar unintended consequences plague (and thereby constitute decisive objections to!) positive intergroup contact, Anderson's preferred strategy. That is, they claim that intergroup contact should be deprioritized precisely because it prevents individuals from recognizing and mobilizing against ongoing injustices. But structural prioritizers repeatedly overstate the practical significance of discovering that certain interventions, when pursued in isolation, come with side effects. If interventions that successfully promote justice also promote complacency,

²⁴ Compare, for example, Shelby's (2014) "egalitarian pluralist" approach to redressing racial inequality.

that just means we should couple those interventions with strategies to resist complacency, and relentlessly remind ourselves how far we still have to go.

ACKNOWLEDGMENTS

I am grateful for feedback on earlier versions of this essay from David Adams, Jacob Alonzo, Saray Ayala, Michael Brownstein, Yareli Carrillo, Michael Cholbi, Brian Dean, Josh Ebner, Katie Gasdaglis, Matthew Goren, Derick Hughes, Victoria Plaut, Peter Ross, Jennifer Saul, Susan Sterrett, Christina Stevens Carbone, anonymous referees at *Ergo*, and from others at Victoria Plaut's Culture, Diversity, and Intergroup Relations Lab in March, 2014; the Bay Area Forum for Law and Ethics in April, 2014; the Brown Bag Research Series at Cal Poly Pomona in November, 2015; my *Philosophy & Psychology of Implicit Bias* class at Cal Poly Pomona in May, 2016; and the *Bias in Context: Psychological and Structural Explanations* conference at the University of Sheffield in September, 2016. This essay was written and revised with institutional support from the Mellon Postdoctoral Fellowship at the University of California, Berkeley; Vassar College; and California State Polytechnic University, Pomona.

REFERENCES

- Alcoff, Linda M. (2010). Epistemic identities. *Episteme*, 7(02), 128-137.
- Alexander, Michelle. (2010/2012). *The New Jim Crow: Mass Incarceration in the Era of Colorblindness*. The New Press.
- Anderson, Elizabeth. (2010). *The Imperative of Integration*. Princeton University Press.
- Anderson, Elizabeth. (2012a). Epistemic justice as a virtue of social institutions. *Social Epistemology*, 26(2), 163-173.
- Anderson, Elizabeth. (2012b). Reply to John Christman's comments. *Social Epistemology Review and Reply Collective* 1 (7): 15-16.

- Antony, Louise. (1995). Sisters, Please, I'd Rather Do It Myself: A Defense of Individualism in Feminist Epistemology. *Philosophical Topics*, 23 (2), 59-94.
- Asgari, Shaki, Nilanjana Dasgupta, and Nicole G. Cote (2010). When Does Contact with Successful Ingroup Members Change Self-Stereotypes?. *Social Psychology*, 41(3), 203-211.
- Asgari, Shaki, Nilanjana Dasgupta, and Jane G. Stout (2012). When do counterstereotypic ingroup members inspire versus deflate? The effect of successful professional women on young women's leadership self-concept. *Personality and Social Psychology Bulletin*, 38(3), 370-383.
- Ayala, Saray and Nadya Vasilyeva. (2015). Explaining Speech Injustice: Individualistic vs. Structural Explanation. *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Cognitive Science Society.
- Banks, Ralph R. and Richard T. Ford (2009). (How) Does Unconscious Bias Matter: Law, Politics, and Racial Inequality. *Emory LJ*, 58, 1053.
- Banks, Ralph R. and Richard T. Ford (2011). Does Unconscious Bias Matter? *Poverty & Race*, Volume 20, Number 5.
- Becker, Dana, and Jeanne Marecek (2008). Dreaming the American dream: Individualism and positive psychology. *Social and Personality Psychology Compass*, 2(5), 1767-1780.
- Brownstein, Michael and Alex Madva. (2012). The normativity of automaticity. *Mind & Language*, 27(4), 410-434.
- Christman, John. (2012). Comments on Elizabeth Anderson, "Epistemic Justice as a Virtue of Social Institutions". *Social Epistemology Review and Reply Collective* 1.
- Dasgupta, Nilanjana and Shaki Asgari. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on automatic gender stereotyping. *Journal of Experimental Social Psychology* 40, 642-658.
- Dasgupta, Nilanjana. (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.
- Dixon, John, Mark Levine, Steve Reicher, and Kevin Durrheim. (2012). Beyond prejudice: Are negative evaluations the problem and is getting us to like one another more the solution?. *Behavioral and Brain Sciences*, 35(6), 411.
- Dotson, Kristi. (2012). A cautionary tale: On limiting epistemic oppression. *Frontiers: A Journal of Women Studies*, 33(1), 24-47.
- Eagly, Alice H., and Steven J. Karau. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109(3), 573.
- Eberl, Carolin, Reinout W. Wiers, Steffen Pawelczack, Mike Rinck, Eni S. Becker, and Johannes Lindenmeyer. (2013). Approach bias modification in alcohol dependence: Do clinical effects replicate and for whom does it work best?. *Developmental cognitive neuroscience*, 4, 38-51.
- Ebert, Irena D., Melanie C. Steffens, Rul Von Stülpnagel, and Petra Jelenec. (2009). How to like yourself better, or chocolate less: Changing implicit attitudes with one IAT task. *Journal of Experimental Social Psychology*, 45(5), 1098-1104.
- Forbes, Chad E., and Toni Schmader. (2010). Retraining attitudes and stereotypes to affect motivation and cognitive capacity under stereotype threat. *Journal of personality and social psychology* 99(5), 740.

- Handley, Ian M., Elizabeth R. Brown, Corinne A. Moss-Racusin, and Jessi L. Smith. (2015). Quality of evidence revealing subtle gender biases in science is in the eye of the beholder. *Proceedings of the National Academy of Sciences*, 112(43), 13201-13206.
- Hannah-Jones, Nikole Chana Joffe-Walt. (2015, August 7). The Problem We All Live With - Part Two [Radio broadcast episode]. In I. Glass (Producer), *This American Life*. Chicago, IL: National Public Radio.
- Haslanger, Sally. (2015a.) What is a (Social) Structural Explanation? *Philosophical Studies* 172.
- Haslanger, Sally. (2015b.) Social Structure, Narrative, and Explanation. *Canadian Journal of Philosophy* 45 (1):1-15.
- Huebner, Bryce. (2016). Implicit Bias, Reinforcement Learning, and Scaffolded Moral Cognition. *Implicit Bias and Philosophy: Volume I: Metaphysics and Epistemology*. Oxford University Press.
- Johnson, India R., and Kentaro Fujita. (2012). Change We Can Believe In: Using Perceptions of Changeability to Promote System-Change Motives Over System-Justification Motives in Information Search. *Psychological science*, 23(2), 133-140.
- Jolls, Christine, and Cass R. Sunstein. The law of implicit bias. *California Law Review* (2006): 969-996.
- Kaiser, Cheryl R., Brenda Major, Ines Jurcevic, Tessa L. Dover, Laura M. Brady, and Jenessa R. Shapiro. (2012). Presumed fair: Ironic effects of organizational diversity structures.
- Kang, Jerry, and Mahzarin R. Banaji. (2006). Fair Measures: A Behavioral Realist Revision of "Affirmative Action". *California Law Review*, 94(4), 1063-1118.
- Kawakami, Kerry, John F. Dovidio, and Simone van Kamp. (2007). The Impact of Counterstereotypic Training and Related Correction Processes on the Application of Stereotypes. *Group Processes and Intergroup Relations* 10 (2), 139-156.
- Kawakami, Kerry, Curtis E. Phillips, Jennifer R. Steele, and John F. Dovidio. (2007). (Close) Distance Makes the Heart Grow Fonder: Improving Implicit Racial Attitudes and Interracial Interactions Through Approach Behaviors. *Journal of Personality and Social Psychology*, 92(6), 957-971.
- Kenyon, Tim. (2014). False polarization: debiasing as applied social epistemology. *Synthese*, 191(11), 2529-2547.
- Kohn, Alfie. (1997). The trouble with character education. In A. Molnar (Ed.), *The Construction of Children's Character. 96th Yearbook of the National Society for the Study of Education, Part II*, 154-62. Chicago: University of Chicago Press.
- Kukla, Rebecca. (2014). Performative force, convention, and discursive injustice. *Hypatia*, 29(2), 440-457.
- Kunda, Ziva, and Steven J. Spencer. When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological bulletin*, 129(4), 522.
- Lord, Charles G., Lee Ross, and Mark R. Lepper. (1979). Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence. *Journal of personality and social psychology*, 37(11), 2098.
- Machery, Edouard, Luc Faucher, and Daniel R. Kelly. (2010). On the alleged inadequacies of psychological explanations of racism. *The Monist; An International Quarterly Journal of General Philosophical Inquiry*, 93(2), 228.

- Madva, Alex. (2016a). Virtue, social knowledge, and implicit bias. Brownstein, M. and Saul, J. (Eds.) *Implicit Bias & Philosophy: Volume 1: Metaphysics and Epistemology*. Oxford: Oxford University Press.
- Madva, Alex. (2016b). Why implicit attitudes are (probably) not beliefs. *Synthese*, 193, 2659–2684.
- Madva, Alex. (ms.). Biased against de-biasing: On the role of (institutionally sponsored) self-transformation in the struggle against prejudice.
- Madva, Alex and Michael Brownstein. (Forthcoming). Stereotypes, Prejudice, and the Taxonomy of the Implicit Social Mind. *Noûs*.
- Markus, Hazel Rose, and Shinobu Kitayama. (2010). Cultures and selves A cycle of mutual constitution. *Perspectives on Psychological Science*, 5(4), 420-430.
- Mann, Nikki H., and Kerry Kawakami. (2012). The long steep path to equality: Progressing on egalitarian goals. *Journal of Experimental Psychology: General*, 141, 187-197.
- Mizell, Jill. (2012). An Overview of Research on Attitudes Toward Expanding Opportunity. Report for The Opportunity Agenda, Tides Center.
- MIT Report. (2011). A Report on the Status of Women Faculty in the Schools of Science and Engineering at MIT.
- Moran, Tal, and Yoav Bar-Anan. (2013). The effect of object–valence relations on automatic evaluation. *Cognition & Emotion*, 27(4), 743-752.
- Moskowitz, Gordon B. (2010). On the control over stereotype activation and stereotype inhibition. *Social and Personality Psychology Compass*, 4(2), 140-158.
- Pettigrew, Thomas F., and Linda R. Tropp. (2006). A meta-analytic test of intergroup contact theory. *Journal of personality and social psychology*, 90(5), 751.
- Purpel, David E. (1997). The politics of character education. In A. Molnar (Ed.), *The Construction of Children's Character. 96th Yearbook of the National Society for the Study of Education, Part II*, 140-53. University of Chicago Press.
- Rydell, Robert J., Allen R. McConnell, Diane M. Mackie, and Laura M. Strain. (2006). Of Two Minds: Forming and Changing Valence-Inconsistent Implicit and Explicit Attitudes. *Psychological Science* 17 (11), 954-958.
- Sarkissian, Hagop. (2010). Minor Tweaks, Major Payoffs: The Problems and Promise of Situationism in Moral Philosophy. *Philosopher's Imprint*, 10(9).
- Shelby, Tommie. (2014). Integration, Inequality, and Imperatives of Justice: A Review Essay. *Philosophy & Public Affairs*, 42(3), 253-285.
- Stewart, Tracie L., Ioana M. Latu, Nyla R. Branscombe, and H. Ted Denney. (2010). Yes we can! Prejudice reduction through seeing (inequality) and believing (in social change). *Psychological Science*.
- Stout, Jane G., Nilanjana Dasgupta, Matthew Hunsinger, and Melissa A. McManus. (2011). STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of personality and social psychology*, 100(2), 255.
- Uhlmann, Eric Luis, Victoria L. Brescoll, and Edouard Machery. (2010). The motives underlying stereotype-based discrimination against members of stigmatized groups. *Social Justice Research*, 23(1), 1-16.
- van Zomeren, Martijn. (2013). Four core social-psychological motivations to undertake collective action. *Social and Personality Psychology Compass*, (6), 378-388.

- Wiers, Reinout W., Carolin Eberl, Mike Rinck, Eni S. Becker, and Johannes Lindenmeyer. (2011). Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychological Science*, 22(4), 490-497.
- Winton, Sue. (2012). Character education: Implications for critical democracy. *International Critical Childhood Policy Studies Journal*, 1(1), 42-63.
- Wright, Stephen C., and Gamze Baray. (2012). Chapter 11: Models of social change in social psychology: collective action or prejudice reduction? Conflict or harmony? In John Dixon and Mark Levine, M. (Eds.). *Beyond prejudice: Extending the social psychology of conflict, inequality and social change*. Cambridge University Press.