

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

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.

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

Although I support many of the structural interventions advocated by Haslanger and others (whom I'll call "Prioritizers"), and although I do not defend political or methodological individualism *per se* in any form, I believe it is false and misleading to claim that we should "prioritize" structural over individual change. My view is not that we should *instead prioritize* individual change, but that individual changes will be integral to the success of structural changes.² I support "anti-prioritization," which emphasizes the interdependence of individual and structural change, although, as I will explain, fighting discrimination effectively will likely require prioritizing *certain* individual changes over others, and prioritizing certain structural changes over others. I submit that Prioritizers' arguments turn on *non sequiturs*, their policy predictions turn on oversimplified empirical models, and their attempt to force a choice between structural and individual reforms is confused.

It is ultimately an empirical question which interventions are most conducive to which social changes. I begin (§2) by assuming that Prioritizers' empirical claims are accurate. My focus is on an internal feature of their arguments. Even if their predictions about effective interventions are basically right, their broader conclusions about the relative priority of individual versus structural reform don't follow. Prioritizers urge a redirection of attention, claiming that we should think less about the individual and more about the social. This is a false choice. They should urge instead that we think *differently* about the individual, and thereby think differently about the social. Subsequently (§3), I contest the Prioritizers' empirical claims. While Prioritizers believe that structural interventions are the best way to realize lasting individual change, I explain how such interventions can be inadequate, short-lived, and even counterproductive when they discount the complexities and contingencies inherent in how individuals interpret and react to changes in the social environment. In this sense, structural

² See, for example, Christman's (2012) comment on Anderson (2012a).

reforms falter by not being “individualist” enough. That is, Prioritizers rely on a problematically oversimplified model of the mind (§4), and a correspondingly misguided model of social change. I conclude (§5), however, by describing what I take to be a kernel of truth underlying their concerns.

II. ARGUMENTS FOR PRIORITIZING STRUCTURAL CHANGE

The individual-level reform on which I focus is prejudice reduction. 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. 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...

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 Prioritizers, focusing on individuals’ prejudiced attitudes “occludes” the primary causes of injustice, which are structural in nature. Consider two examples.

First, consider structural factors contributing to racial inequality. Elizabeth Anderson (2012a, 171), for instance, 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 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 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, 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 “respectable” or hard-working. In fact, 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 §§4-5). 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).

Second, consider a structural factor contributing to gender inequality. Disparities between maternal and paternal leave policies encourage mothers to leave the work force more than fathers, and thereby reinforce patterns where men earn more, have more opportunity for career advancement, have more domestic power, and so on. Again, parental leaves policies are features of the systems in which individuals make decisions: they make certain actions possible

and others impossible, and they make some options much more feasible and attractive than others. Such policies can reinforce, or even create, an unjust state of affairs even if all the individual actors—fathers, mothers, employers—are acting rationally; are uninfluenced by gender biases or stereotypes; and start out with identical career goals, caregiving preferences, political commitments, and so on. Prioritizers infer that a specific type of structural change—in this case, reforming parental leave policies—should be prioritized over specific types of individual change, such as reducing individuals’ (conscious or unconscious) biases against working mothers and caregiving fathers, or strengthening individuals’ commitments to gender egalitarianism.

Suppose, in this second example, that it is true that these particular policy changes should be prioritized over these particular individual reforms. Does it follow that we should, in this case, prioritize structural reforms in general over individual reforms in general? I don’t see how. *Bringing about* these policy changes requires, at a minimum, changes in the beliefs, motivations, or actions of those individuals poised to help change policy.³ 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 (a point I return to in §3). 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 effecting a necessary

³ Of course, the relevant structural changes *could* occur without anybody intending to bring them about, e.g., as an unintended benefit from some other change, but Prioritizers are not pinning their hopes on such accidents. We are restricting attention to possible worlds in which some intentional individual or collective efforts are the drivers of change. Also, note that individual changes can be necessary without being sufficient for structural change. To view individual change as sufficient would be to endorse a kind of methodological individualism. But here my claim is that even if methodological individualism is false, that doesn’t make Prioritization true.

structural reform, and therefore to promoting justice, than is changing *other* individual attitudes (e.g., individuals' motivation to reform parental leave policies). This example also requires us to assume that individuals' biases about gender, careers, and caregiving exert no meaningful influence on their beliefs and motivations surrounding the reform of parental leave policies, which seems implausible. Changing individuals' gender biases might very well be integral to drumming up support for reforming parental leave.

Parallel considerations apply to the first example. Suppose Anderson is right that “segregation is the structural ground” of inequality and that the foremost “structural remedy” is integration (2012a, 171). 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 any independent attempts to reform individuals' prejudices, such as diversity training. But it does not follow that we should prioritize structural over individual reform full stop. The most Anderson's defense of integration can show, if it were correct, is that certain structural reforms should be prioritized over others, and certain individual reforms should be prioritized 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-black communities.⁴ Regarding individual reforms, perhaps motivating individuals to support integration should take precedence over reducing their prejudices. As in the previous example, for this latter contrast between types of individual reform to make sense, we would have to

⁴ 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.

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 racial 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.

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 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.⁵ This is not the model. Positive

⁵ See, for example, how Wright and Baray (2012) contrast what they call the “causal flow” through three “levels of analysis” of individualist versus anti-individualist models, where the “micro-level” of analysis refers to “intrapersonal phenomena (cognitions, emotions etc. within individual people),” the “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” (236).

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 mysteriously elide these structural interventions in their portrayals of the contact hypothesis. In fact, contact theorists—such as Anderson—tend to be 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. Suppose that reducing individuals' prejudices through intergroup contact does not advance social justice, and even hampers it. Then researchers were wrong to focus on this particular type of individual change, namely, prejudice reduction.⁶ By itself, such a finding wouldn't warrant a shift of attention away from strategies for promoting individual change, but a shift of attention to strategies for promoting different types of individual change. 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 includes institutional, structural, and social factors, for example, a national voting holiday, accessible public transportation to voting stations, and peer pressure to march and vote—and also

⁶ Perhaps researchers also focused on the wrong individuals, i.e., the advantaged instead of the disadvantaged. Certain types of negative attitude toward the advantaged may actually promote activism among the disadvantaged.

psychological factors, like the desire to effect change and the belief that marching and voting will be instrumental to doing so.

Indeed, when 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).⁷ 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 (i.e., making them more likely to focus on the bad stuff that needs to be changed) and also increases their motivation to change the system (see also Stewart et al. 2010). Perhaps changing individuals’ attitudes about the malleability of social structures is more important than changing their attitudes about other social groups.

In a companion paper (under review⁸), 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, 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

⁷ They 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.

⁸ “Biased Against Debiasing: On the Role of (Institutionally Sponsored) Self-Transformation in the Struggle Against Prejudice” (under review). A draft of the paper is available online [here](#).

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), increase self-esteem (Ebert et al. 2009), 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), and change preferences for *Haribo* gummy bears versus *Milka* chocolate bars (Ebert et al. 2009). 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 Prioritizers give us no reason to think that these psychological changes are somehow irrelevant to that end. *If* they are right that prejudice reduction *per se* should not be prioritized, the upshot is not a general lesson about the pitfalls of pursuing individual changes in the absence of structural changes. The upshot is just that prejudice reduction is the wrong individual change to prioritize, as opposed to individual changes toward increased political participation.

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

In fact, Anderson at times nearly claims that racial justice can be promoted without individuals fighting for it, for example, “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. Subordinates 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 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 white supremacists), and it is not always easy to come by (for example, when authority figures are perceived as authoritarian). 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 de-prioritize individual change in general, but to de-prioritize some individual changes and prioritize others.

III. STRUCTURAL CHANGE AS PRECONDITION FOR INDIVIDUAL CHANGE

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. 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 (Anderson 2010, 2012a,b; Dasgupta 2013; Huebner forthcoming; Jolls and Sunstein 2006; Kang and Banaji 2006). 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 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? 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; Stout et al. 2011). However, the centerpiece of Prioritizers' defense of affirmative action as a two-bird stone is that it will change the biases of members of *other* social groups (e.g., Kang and Banaji 2006, 1109-10), but research increasingly suggests that having women professors has *no effect whatsoever* on undergraduate

men's stereotypes about math ability and gender. This particular debiasing effect applies only to ingroup members (women) rather than outgroup members (men).⁹ 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 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.¹⁰ 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

⁹ 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 role-modeling defense of affirmative action in large part because it has been shot down by the courts.

¹⁰ See, for example, Eagly and Karau (2002) and Valian (1998). See Kukla (2014) for insightful discussion of similar cases.

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 “break the public association of blacks with poverty and associated dysfunctional behaviors,” the overall consequences of such a revision in public perception are likely complex. Consider some of Michelle Alexander’s (2010/2012, 248) 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.

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 prevalent claim that the election of a black President signals that we are in 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.¹¹

As further evidence for backlash effects of structural reforms on individual attitudes, take the recent case of MIT. 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. 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 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 “take care of themselves” or gradually fade with time.¹² They may instead grow, and eventually

¹¹ 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 my “Virtue, Social Knowledge, and Implicit Bias” (forthcoming) for further discussion. A draft is available online [here](#).

¹² 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 Prioritization. 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

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” (Alexander 2010/2012).

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 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. Discrimination and inequality are *two-stone birds* that must be jointly tackled by individual and structural reforms.¹³

To be fair to Prioritizers, there are many studies that do show that exposure to counterstereotypes can reduce individuals’ biases. I say more about these studies elsewhere (Madva under review), 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,

predictable, and therefore poses an obvious problem for Prioritizers. 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 know full well that widespread resistance to the intervention makes it “politically unfeasible” (189).

¹³ 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*.

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 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.¹⁴ This active role is elided in 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. 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 minds, according to Prioritizers.¹⁵ Call this view MIRROR. Specifically, to say that our biases are

¹⁴ See Sarkissian (2010) for insightful discussion of similar pitfalls in the “situationist” critique of virtue ethics.

¹⁵ See, for example, Huebner (forthcoming) 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

“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 group, 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 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 (forthcoming) writes:

So as we watch or read the news, watch films, rely on tacit assumptions about what is

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 the specific anti-individualist models of *how* social phenomena influence psychological phenomena.

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 for Prioritizers. 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.

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.¹⁶ 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.¹⁷ 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

¹⁶ This paragraph and the next draw in part from Madva (under review).

¹⁷ See Kenyon (2014) for discussion of polarization biases and arguments for prioritizing social over individual reforms to address these biases.

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. Far from it. 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; see Madva forthcoming, under review for further discussion). 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 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, Prioritizers including Anderson (2010, 2012a), Haslanger (2015b), and Huebner (forthcoming) 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 nothing to do with individualism.

**V. THE KERNEL OF TRUTH BEHIND 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. I think the 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.¹⁸

One way of bringing this out is to return to debates among Prioritizers regarding which structural interventions to prioritize (§2). Anderson argues that integration is the structural intervention *par excellence* both for redressing inequality and for reducing prejudice. Other Prioritizers, 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 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 *motivates* this criticism warrants serious attention. This is the kind of evidence I discussed in §2 of “ironic,” 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

¹⁸ Those whom I call Prioritizers have of course produced many valuable insights. I’m not saying there’s only one kernel of truth in their oeuvre. I’m speaking specifically about the claim to prioritize structural over individual reform.

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 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?¹⁹ 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, given limited resources, we cannot feasibly both give reparations and promote integration. Although I am skeptical about the applicability of an argument from limited resources in this context, this response is plainly unavailable to Prioritizers, who argue that many different and intensive structural reforms are needed to

¹⁹ Compare, for example, Shelby's (2014) "egalitarian pluralist" approach to redressing racial inequality.

effectively combat discrimination and inequality. “Integration” 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. 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, that just means we should couple those interventions with strategies to resist complacency, and relentlessly remind ourselves how far we still have to go.

REFERENCES

- Alcoff, L. M. (2010). Epistemic identities. *Episteme*, 7(02), 128-137.
- Alexander, M. (2010/2012). *The New Jim Crow: Mass Incarceration in the Era of Colorblindness*. The New Press.
- Anderson, E. (2010). *The Imperative of Integration*. Princeton University Press.
- Anderson, E. (2012a). Epistemic justice as a virtue of social institutions. *Social Epistemology*, 26(2), 163-173.
- Anderson, E. (2012b). Reply to John Christman’s comments. *Social Epistemology Review and Reply Collective* 1 (7): 15-16.
- Asgari, S., Dasgupta, N., & Cote, N. G. (2010). When Does Contact with Successful Ingroup Members Change Self-Stereotypes?. *Social Psychology*, 41(3), 203-211.
- Asgari, S., Dasgupta, N., & Stout, J. G. (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.
- Banks, R. R., & Ford, R. T. (2009). (How) Does Unconscious Bias Matter: Law, Politics, and Racial Inequality. *Emory LJ*, 58, 1053.
- Banks, R.R. & Ford, R.T. (2011). Does Unconscious Bias Matter? *Poverty & Race*, Volume 20, Number 5.
- Christman, J. (2012). Comments on Elizabeth Anderson, “Epistemic Justice as a Virtue of Social Institutions”. *Social Epistemology Review and Reply Collective* 1.

- Dasgupta, N., and 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.
- 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.
- Dixon, J., Levine, M., Reicher, S., & 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.
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109(3), 573.
- Eberl, C., Wiers, R. W., Pawelczack, S., Rinck, M., Becker, E. S., & Lindenmeyer, 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.
- Ebert, I. D., Steffens, M. C., Von Stülpnagel, R., & Jelenec, P. (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, C. E., & Schmader, T. (2010). Retraining attitudes and stereotypes to affect motivation and cognitive capacity under stereotype threat. *Journal of personality and social psychology* 99(5), 740.
- Handley, I. M., Brown, E. R., Moss-Racusin, C. A., & Smith, J. L. (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, N., & Joffe-Walt, C. (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, S. 2015a. What is a (Social) Structural Explanation? *Philosophical Studies* 172.
- Haslanger, S. 2015b. Social Structure, Narrative, and Explanation. *Canadian Journal of Philosophy* 45 (1):1-15.
- Huebner, B. Forthcoming. Implicit Bias, Reinforcement Learning, and Scaffolded Moral Cognition. *Implicit Bias and Philosophy: Volume I: Metaphysics and Epistemology*. Oxford University Press.
- Johnson, I. R., & Fujita, K. (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, C. & Sunstein, C. R.. The law of implicit bias. *California Law Review* (2006): 969-996.
- Kaiser, C. R., Major, B., Jurcevic, I., Dover, T. L., Brady, L. M., & Shapiro, J. R. (2012). Presumed fair: Ironic effects of organizational diversity structures.
- Kang, J., & Banaji, M. R. (2006). Fair Measures: A Behavioral Realist Revision of "Affirmative Action". *California Law Review*, 94(4), 1063-1118.
- Kawakami, K., Dovidio, J.F., and van Kamp, S. 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, K., Phills, C.E., Steele, J.R., and Dovidio, J.F. 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, T. (2014). False polarization: debiasing as applied social epistemology. *Synthese*, 191(11), 2529-2547.
- Kukla, R. (2014). Performative force, convention, and discursive injustice. *Hypatia*, 29(2), 440-457.
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence. *Journal of personality and social psychology*, 37(11), 2098.
- Mann, N. H., & Kawakami, K. (2012). The long steep path to equality: Progressing on egalitarian goals. *Journal of Experimental Psychology: General*, 141, 187-197.
- Mizell, J. (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.
- Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of personality and social psychology*, 90(5), 751.
- Sarkissian, H. (2010). Minor Tweaks, Major Payoffs: The Problems and Promise of Situationism in Moral Philosophy. *Philosopher's Imprint*, 10(9).
- Shelby, T. (2014). Integration, Inequality, and Imperatives of Justice: A Review Essay. *Philosophy & Public Affairs*, 42(3), 253-285.
- Stewart, T. L., Latu, I. M., Branscombe, N. R., & Denney, H. T. (2010). Yes we can! Prejudice reduction through seeing (inequality) and believing (in social change). *Psychological Science*.
- Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (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.
- Wiers, R. W., Eberl, C., Rinck, M., Becker, E. S., & Lindenmeyer, J. (2011). Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychological Science*, 22(4), 490-497.
- Wright, S. C., & Baray, G. (2012). Chapter 11: Models of social change in social psychology: collective action or prejudice reduction? Conflict or harmony? In J. Dixon & M. Levine, M. (Eds.). *Beyond prejudice: Extending the social psychology of conflict, inequality and social change*. Cambridge University Press.