

First let me say that the standard line about “being honored” to receive these comments would be a profound understatement. Getting this kind of feedback from three thinkers I admire so much—Saray Ayala, Sally Haslanger, and Jenny Saul—is possibly the coolest thing to happen in my professional career. So I’m also very grateful to John Schwenkler, Aaron Henry, and the Brains Blog for making this happen!

1. The elephant in the room

Although my paper is neck-deep in empirical research and very applied questions, some of its central aims are conceptual: to draw attention to inferential missteps that, I believe, crop up frequently in social-justice-oriented discussions, including in academic philosophy, political activism, journalism, the social sciences, etc. My hope is that, like many conceptual oversights, once they’re pointed out, they seem obvious. As a result, my paper sometimes tiptoes along the tightrope between the trivially true and patently false. So I’ll start by trying to tackle the elephant in the room: am I being uncharitable to those I call structural prioritizers, because nobody actually believes the claims or draws the inferences that I attribute to them? Since it seems clear that individual and structural reforms are interdependent, the idea that we should prioritize one over another doesn’t make sense, and perhaps it’s uncharitable to suggest that so-called prioritizers think otherwise.

Specifically, Ayala, Haslanger, and others such as Ruth Groff have suggested that I don’t properly contextualize structural prioritizers’ claims, in at least two respects. First, perhaps structural-prioritist claims should not be read *strategically* (as regarding the means) but *normatively* (as regarding the final ends, which Ayala calls “goal states”), i.e., not actually about what we should try to do here and now but what we should ultimately be fighting for. Our primary aim should not (merely?) be the elimination of individuals’ bad attitudes but the

transformation of unjust structures, to eliminate underlying inequalities. Second, Ayala and Haslanger suggest that prioritizers should be read as reacting specifically to the recent uptick of interest in implicit bias. However, I don't think that, as a rule, structural prioritizers contextualize their claims in either of these ways. Regarding the first, although I'd like to reiterate that I'm deeply sympathetic with prioritizers' basic normative analyses of racial injustice and justice, their normative discussions often shift seamlessly into very concrete empirical predictions and strategic means-ends prescriptions, which is what I take issue with in my paper. Regarding the second, people make structural-prioritist claims *all the time* in all sorts of contexts, often in an offhand, slogan-like way (as if the claim is so obvious it doesn't require elaboration or defense). An *Ergo* referee drew my attention, for example, to parallel debates in character education and positive psychology, and similar debates are clearly at play elsewhere, e.g., whether trying to persuade individuals to "reduce, reuse, and recycle" is a waste of time, in contrast to more profound structural reforms to protect the environment.

Consider the following paragraph from Haslanger's "Oppressions: Racial and Other," which was first published in 2004, and is not about implicit bias. It seems to me that this passage begins with a definitive acknowledgment of individual-structural interdependence, and then transitions into an unqualified insistence on structural prioritism (*Resisting Reality*, ch.11, p.319, all emphases mine):

Of course both kinds of situation certainly obtain and are of concern: our societies are unjustly structured, and immoral people with power can and do harm others. Moreover, individual and structural issues are interdependent insofar as individuals are responsive to their social context and social structures are created, maintained, and transformed by individuals. Nonetheless, there will be situations that are clearly unjust even when it is unclear whether there is an agent responsible for the oppression; we don't need a smoking gun to tell that a system of practices and policies that result in women being denied adequate health care is unjust. I also submit that we should have more hope in the prospects of social and political change bringing about a significant improvement in people's lives than

in the prospects of anything like the moral improvement of individuals. As Liam Murphy (1999: 252) suggests: “it is obviously true that, as a practical matter, it is overwhelmingly preferable that justice be promoted through institutional reform rather than through the uncoordinated efforts of individuals—a point worth emphasizing in an era characterized by the state’s abandonment of its responsibility to secure even minimal economic justice and by politicians’ embrace of ‘volunteerism’ as a supposed substitute.” I will not attempt to justify this hope in structural as opposed to moral reform here. But in my experience, not only is structural reform usually more sweeping and reliable, but it also allows ordinary individuals who unwittingly contribute to injustice to recognize this and change their ways, without the kind of defensiveness that emerges when they find themselves the subject of moral reproach.

Let me begin by stressing that *of course* I know roughly what Haslanger is getting at here, and I’m very sympathetic with much of it. For example, all parties in this conversation agree that structural change is necessary (a view Ayala calls “structural necessarism”), and relatedly, that individual change is insufficient. That said, this paragraph contains several of the sorts of claims and inferences of concern in my paper.

First, structural reform is here characterized as better *in general*—without having to specify *which* structural reforms are better, or why, or how—on the basis of evidence that *certain specific* forms of individual reform are not effective (in this case, the bogeyperson is uncoordinated volunteerism, i.e., individual charity and “self-taxing”¹). Thus the inefficacy of the *wrong kind* of individual change is used to infer that individual change is worse than structural change full stop.

Second, Haslanger claims as decisive advantages for structural reforms that they are more sweeping and reliable, and, by avoiding defensiveness, facilitate the necessary sorts of individual transformation (as I explain in my paper, structural reforms are described as killing two birds with one stone). But as she rightly insists in her comments here, such claims are mistaken:

¹ Cf. Anderson’s comparison of Fricker’s account of testimonial virtue to individuals donating to charities (Anderson, 2012, p. 171).

certain structural reforms—including *poorly implemented* structural reforms—can elicit a great deal of backlash. The tendency to see structural reforms as more reliable may reflect a kind of anti-individualist confirmation bias. When we say things like this, we’re simply not thinking of high-profile failed structural transformations, such as the passage of a Constitutional amendment to outlaw alcohol consumption, which resulted from a long-fought collective-action campaign and seemed at the time to be as sweeping and reliable as any reform could be. A currently salient example is the Affordable Care Act, which surely seems less “locked in” now than it did just a few months ago. In fact, the ACA was one or two Supreme Court votes away from being struck down on several occasions. And if it does manage to survive the current administration relatively unscathed, we should be cautious about interpreting this result as further evidence for the claim that structural change tends to be more stable. Psychological factors, such as loss aversion, and idiosyncratic features of relevantly-positioned individuals, such as the incompetence of the current administration, are certainly contributing to the difficulty of repeal. More generally, whether a given structural reform “sticks” does not depend solely on the reform itself but on a host of other factors, including its psychological uptake. Some structures become like water to the fish, while others remain intrusive, obtrusive, or are even experienced as oppressive.

Third, while Haslanger does not elaborate on what she has in mind by “anything like the moral improvement of individuals,” it’s a manifestly narrow construal of moral improvement. It seems to include uncoordinated volunteerism and probably prejudice reduction; she’s alluding to individuals who treat other individuals fairly (at least according to the existing rules) but continue to operate in unjust systems. But consider an individual who goes from not believing to believing that *she should work to bring about structural reform*. That’s a change in individuals attitudes, and (I would say) a moral improvement, and also a plausible candidate for the kind of

individual-level reform that, without being necessary or sufficient for bringing about justice, will very likely be conducive to it. I gather Haslanger would agree, in which case the question is not whether to change attitudes vs. structures, but which attitudes and structures, and how. On the individualistic side, a crucial question becomes how can we motivate others and ourselves to form and live up to such moral commitments, and develop the affective-cognitive-bodily habits to embody these commitments in our daily lives. This leads me to conclude that the slogan to redirect attention from the individual to the social is not just false but seriously misleading, because it can prevent us from thinking more carefully about which aspects of individual hearts and minds to change in order to bring about the structural reforms that are necessary to achieving our final ends (or, more modestly, to moving toward the transition states along the way to our final ends, as Ayala discusses). Our slogans should call instead for a *binocular focus* on individuals-in-structures and structures-around-individuals (or a poly-ocular focus, since there are plausibly many more than two dimensions to consider—although perhaps any appeal to ocular metaphors is problematic).

Fourth, there's the question how to square Haslanger's auspicious nod in this passage to individual-structural interdependence with her subsequent endorsement of structural prioritism. I'm struck that her assertion of interdependence occurs in a broadly backward-looking context, of explaining injustice, while the assertion of prioritism occurs in a more forward-looking context, regarding what to do. So my diagnosis—and I think activists, policymakers, journalists, and academics across numerous disciplines make roughly this move all the time—is that folks tend to be much more sensitive to the relevant complexities and interdependencies in examinations of the perpetuation of injustice (or in after-the-fact discussions of why some promising intervention failed, or incurred unintended consequences), but then when we switch

gears to theorizing paths to justice, the complexity often drops out. Again, the ACA is a plausible example. Apparently, the Obama administration predicted that initial resistance would simply fade away all by itself once enough people started to reap the benefits. Instead, there were widespread (perceptions of) rising costs, the “undeserving” getting handouts, etc. Moreover, as is explained in the paper by [Valentino and colleagues](#) to which Saul’s commentary links, racism itself helps to explain anti-ACA backlash. In my paper, I suggest that the view of the mind as a MIRROR of social reality is often partly to blame for such failures of foresight. Admittedly, few would unreservedly endorse MIRROR upon reflection, and Haslanger, for one, absolutely clearly rejects it in many passages. Yet I suspect that MIRROR perseveres at quasi-conscious levels, because it seems especially likely to rear its head in forward-looking contexts. Another way to put this point would be to say that folks like Haslanger are *not doing justice to some of their own best insights* about individual-structural looping and interdependence when they transition from explaining injustice to thinking through the pivot toward justice.

There’s actually a question to raise about who’s being uncharitable to whom, because pretty much every person I know who cares about implicit bias *also* thinks that profound structural reforms are essential. To my mind, one irony here is that many (perhaps most) philosophers and social psychologists who work on implicit bias—such as those I cite in my paper—are *structural prioritizers, too*, with all the problematic implications I’ve just rehearsed. Folks who are justifiably frustrated by the individualistic focus in other spheres see individualism reappearing in discussions of implicit bias, whereas I often incline toward precisely the opposite criticism.² As I explain in [a forthcoming companion paper](#), theorists and

² Arguably, one sense in which the criticism of the philosophy of implicit bias as individualistic is deserved regards the disproportionate attention that’s been given to the question of individual moral responsibility (I am also guilty on this score). That said, much philosophical writing on moral responsibility for implicit bias has discussed social-

activists (usually under the tacit or explicit influence of MIRROR) are far too pessimistic about the prospects of profound, durable individual-level transformation, and so their attention is unduly restricted to larger-scale structural and environmental reforms. On my view, many of these folks are not individualist enough, in that they fall back on an oversimplified view of how the mind interprets and reacts to structural change, and that they fail to take seriously the numerous integral roles that individual changes can play in bringing about and maintaining structural reform.

2. Two more responses to Haslanger: backward-looking analyses and the knowledge-bias relationship

Let me touch on two other points that come up in Haslanger's comments.

First, I did not intend in any way to denigrate the role of backward-looking analysis. I wholeheartedly agree that understanding the explanations, causes, and normative cores of group-based injustice is vitally important, both for their own sake and also for the practical aim of figuring out what to do. Many feminists, critical race theorists, and social scientists (including my commentators) have done tremendous work in this area. In this paper, I intended to focus on the forward-looking question what to do to bring about change. However, a referee helpfully pointed out that the first version I submitted to *Ergo* confusingly alternated back and forth between backward-looking and forward-looking questions. So I added a paragraph to clarify why I talk about backward-looking questions at all, with a principal reason being that structural prioritizers draw the two closely together (and presumably rightly so!). I only meant to note that

structural causes of bias and deficiencies in common knowledge (as excusing factors) and issues like “ecological control” and structural responsibility (as non-excusing factors).

we cannot simply *deduce* proactive strategies from retrospective explanations, and did not intend to suggest that structural prioritizers think otherwise.

Second, Haslanger suggests that part of what warrants pessimism about the prospects of individual-level change is that many of the problematic attitudes are actually epistemically justified and true, e.g., women do tend to be more submissive than men. These attitudes are hard to change because they are (to some extent rationally) reinforced by (unjust) social reality, and they reinforce this reality in turn. I am [on record](#) as arguing that a proper appreciation of the relations between social knowledge, bias, and individuals' (un)ethical behavior speaks against this form of pessimism (I return to this point in footnote 17 of my paper and at the end of Section 5 of the companion piece). Merely knowing that women tend to be more submissive than men is not the problem. Indeed, we *need* to know about such group differences because they reflect real-world injustices to be combatted. Problems depend a) on when and how these bits of knowledge come to mind (e.g., when they are unwittingly activated in irrelevant contexts), and b) on how we are inclined to explain them (e.g., when we make the decidedly *epistemic* mistake of appealing to innate differences rather than social history). Much of the individual-level research of interest to me is oriented towards *regulating* the accessibility of our social knowledge, such that it comes to mind when and only when it is relevant. The kinds of individual transformation I focus on do not threaten our epistemic rationality.

In a related vein, Ayala points out that we don't need to posit MIRROR in order to claim that structural factors play an important role in causing our biases. I agree (Madva forthcoming, especially Section 4). However, Ayala seems to suggest that a less extreme view than MIRROR can also serve to justify something in the ballpark of structural prioritism. But I don't see how one can consistently take a more nuanced and qualified view regarding the role that structures

play in causing biases and still defend anything like prioritism.³ Once we grant that our biases are only *partly* reflections of unjust social reality, and if we are assuming that our biases a) directly contribute to discriminatory behavior as well as b) function to reinforce unjust social reality, then it seems to follow that we will need to tackle both the unjust-social-realities “part” of the story as well as the “other parts,” whatever they are. To take an extremely oversimplified case, suppose that intergroup biases are 50% caused by structural factors and 50% caused by individual-psychological factors (such as the desire to boost one’s self-esteem by putting down other groups). Now suppose that we discover an intervention that makes a significant dent in the individual-psychological causes (e.g., teaching people alternative sources of self-esteem). I come along and say, “it looks like we should teach people alternative sources of self-esteem!” The prioritizer replies, “but this will not fully uproot the structural causes of bias, or all of the underlying inequalities.” But in this scenario, *and unless you endorse MIRROR in some extreme form*, it’s equally true that overhauling the social structures will fail to fully uproot the individual-level causes of bias. Given that the individual-level causes themselves contribute to discrimination and function to reinforce unjust social realities, then they will have to be addressed on their own terms, if not before the structural revolution, then after—lest they fester.

3. Responses to Ayala

The distinctions that Ayala introduces—between CAUSE and EFFECTIVENESS, ORIGIN and MAINTENANCE, structural necessarism and structural sufficientism, transition state and goal state, and then the various competing conceptions of goal states—are enormously helpful and

³ I’ve benefited a great deal in thinking through this issue from discussions with Derick Hughes, and from reading his manuscript, “Rational Autonomy and Irrational Bias: A Reply to Elizabeth Anderson’s Justification for Structural Reform”

important. Our views may depart somewhat regarding what follows once we start taking these distinctions into account, but these matters will certainly require further thought on my part.

For example, Ayala highlights a crucial distinction between originating causal factors (of bias, say) and maintaining causal factors. Perhaps unjust structures play a powerful role in initially causing our biases but play less of a role in maintaining them, such that our biases would persist even after the originating-structural causes are removed. In that case, after the relevant structural intervention, at least unblemished minds might fail to form biases. This kind of outcome is empirically possible (and I know it's just an example and I'm not sure how seriously Ayala takes it), but I find it unlikely, depending on how much we build into the relevant structural intervention. As far as developmental origins go, the idea that the right structural intervention could, on its own, prevent individuals from forming biases in the first place sounds a lot like MIRROR to me. The psychological motivations to draw ingroup-outgroup boundaries, to derogate outgroup members, to consolidate power among the ingroup, etc., are immensely powerful, and I think they would exert an influence even if the basic (non-attitudinal) structures of society were to become just. Indeed, I suspect that these individual-psychological factors have, throughout history, played a crucial role in the ORIGIN and MAINTENANCE of unjust social structures. This means that the structural interventions necessary to prevent minds from forming biases will have to be more targeted, to take these dispositions into account and reduce/prevent the formation of bias. But if we're going to need a bias-targeting component in our structural reform anyway, then why should we abandon the folks (like me) who already have biased minds? Much of the individual-level research I discuss emphatically demonstrates that we old dogs can learn new tricks, so I see no reason to rest satisfied with interventions that fail to reduce already-existing biases. This is especially so because those with blemished minds will

presumably (intentionally or unintentionally) transmit some of their biases to the unblemished minds. That is, the ORIGIN story of bias will surely include an important role for interactions with other biased individuals.

This brings me to another of Ayala's essential distinctions, between transition states and goal states. Ayala suggests, for example, that the integrationist reforms advocated by Anderson might be conceived as *transition* states that reduce some inequalities and facilitate movement toward a fuller social justice. First, it bears mentioning that Anderson often depicts integration as an end in itself (partly constitutive of justice *per se*) and not as a merely instrumentally valuable transition state. (Also, I read Anderson's relational conception to, roughly speaking, *identify* the elimination of certain sorts of inequality with social justice. And she is repeatedly explicit that integration is the ur-two-bird-stone that both reduces inequality and bias, so it's fair to evaluate her proposals with respect to either or both aims.) But I am broadly sympathetic with many of Anderson's integrationist proposals (in fact my paper defends contact theorists like her from the left-based criticism of being too individualist). What I argue is, given that integration is worth pursuing, the straightforwardly empirical question becomes how to promote it, and this will *inevitably* include some individual-level interventions. Many of the interventions I focus on hold promise precisely as ways of *improving* intergroup interaction (Madva forthcoming, Sections 3 and 7), to reduce the kinds of backlash and unintended consequences associated with it. To put this point in Ayala's terms, these sorts of individual-level interventions may help to determine whether integrationist reforms constitute genuine transitions on the road to our ultimate aims, or whether they will be undone and counterproductive (e.g., by leading to white flight, by motivating efforts to undo the structural reforms, etc.). In the "Reconstruction" period after the Civil War, political activism and literacy in the black community skyrocketed. One would have

thought it was a transition state to racial equality. One would have been wrong. I'm sure I'm sounding like a broken record, but my concern is that failing to tackle individuals' motivations and biases will render inequality-reducing structural reforms short-lived.

Ayala suggests that a final reason to be wary of bias-reducing efforts is that what is a bias in one context might not be a bias in another. I agree about the context-specificity of bias (Madva 2016a; Madva forthcoming, Sections 2 and 5). This point would count against any debiasing intervention that would be so totalizing and permanent that it would remain effective across all contexts and could not be unlearned were social reality to subsequently change. We might think here of when the protagonist in *A Clockwork Orange* becomes incapable of committing violence, even in conditions where self-defense might be justified. But I know of no interventions that are so totalizing. Actually, one of the standards knocks on individual-level debiasing techniques is that they're *too* context-specific, so I'm usually in the position of trying to argue that they'll be at least somewhat context-general. But my basic reply here is simply that if social reality changes, and new groups (or social positions) become oppressed or oppressors, then we will have to identify new interventions to fight the new biases. Perhaps we can just change the stimuli in preexisting debiasing techniques, or perhaps more ingenuity will be required. More broadly, I would argue that one of the individual virtues we ought to cultivate is the ability to recognize the contexts in which particular attitudes and actions are biased versus unbiased. As a conceptual point, so long as we are practically reasoning about which strategies to prioritize, there is no getting around the need for individual-level interventions, whether they are explicitly flagged as such or not. For example, if we eliminated all social biases but still lived in a world with problematic presupposition introductions, then presumably one of the things we would want to

do would be to teach individuals strategies for identifying, calling out, and resisting problematic presupposition introductions.

4. Responses to Saul

Speaking of new social realities, Saul worries that my paper, and much recent work on prejudice, discrimination, and inequality, now seems somewhat obsolete. Should we say, with apologies to Virginia Woolf, that on or about November 2016, human nature changed? Or that we just got human nature totally wrong? I see where Saul's coming from, but perhaps predictably, I don't agree. If anything, I think the general concerns I raised in this paper about backlash are looking surprisingly prescient, much more so than I anticipated.

It's true that this paper was first conceived during a very different moment in the "political cycle," when a number of progressive accomplishments were salient in the collective imagination—the election and reelection of the first black President, marriage equality, healthcare reform, the Arab Spring, etc. It seemed to me that the primary individual-level intervention we should have been targeting at that time was how to animate and maintain activist gusto despite these symbolic and substantive achievements. I continue to think that this is a profoundly important empirical question. Assuming that we're not headed toward the total dissolution of democratic elections or nuclear holocaust, then someday the pendulum will swing again, and after 4 or 8 or 2 or however many years of devoting so much energy to fighting in the resistance, then people on the left will again be ready to sit back, relax, and see what's on Netflix again. We have to find ways to maintain at least some degree of activist engagement in the wake of salient signals of progress.

Count me among the folks who found Trump's election basically unthinkable. That said, the existing literature on intergroup psychology is, in fact, extremely well-poised to explain

what's happened. The failure resides in our inability to accept what the data's been telling us all along. (I've added a few slides on what follows to my standard "bias training" shtick.) First, take studies like (Kawakami, Dunn, Karmali, & Dovidio, 2009), which found (a) that most people predicted they would react negatively to discriminatory actions, but (b) that most people actually tended not to react at all when others used an extreme racial slur. Now, (a) helps to explain our confidence that racism would not be tolerated, while (b) explains why our confidence was mistaken! *We should have known* that toleration of such overt bigotry was seriously possible. Second, existing theories about the relations between implicit bias, explicit bias, and behavior straightforwardly predict what's gone down. For one thing, it's long been clear that implicit biases are not unconscious in a deep or distinctive sense. Children, for example, report *explicit* biases very early on. What they learn as they age is that they are not allowed to say them out loud. Roughly speaking:

explicit bias + anti-prejudiced upbringing = implicit bias [i.e., aversive racism]

In turn, the existing evidence also suggests that the following is roughly true:

implicit bias + normalization of prejudice = explicit bias

Predictions like this fall *directly* out of the leading models of intergroup cognition, such as MODE. For example, in (Cooley, Payne, Loersch, & Lei, 2015; Cooley, Payne, & Phillips, 2014), participants first completed a measure of their implicit attitudes toward gay vs. straight couples. Some were then told, "You may have had a 'gut feeling' towards the pictures of

heterosexuals and homosexuals. Research has found that this gut feeling usually reflects people's genuine attitude towards homosexuality." Participants who had strong anti-gay implicit biases were subsequently more likely to report being opposed to gay marriage and military enrollment. By contrast, participants who were told that their gut feelings did *not* reflect their genuine attitudes showed the opposite pattern. Such studies are a clear reminder of the power of authority figures (in this case, scientists, but presumably also professors, parents, and political leaders) to legitimize prejudice and discrimination. However, it must be stressed that a *purely* social-power or norm-based explanation here is inappropriate. This manipulation only affected individuals who had strong anti-gay implicit biases. Those who lacked such biases were unaffected by this prejudice-legitimizing rhetorical maneuver.

The lesson, then, is not that we are mere puppets who ask "how high?" when the right authority figure tells us to jump. Nor is it that our prior emphasis on implicit bias was misplaced. Implicit biases play a fundamental role as the "raw material" out of which explicit biases are, in certain social contexts, manifest, or at least tolerated. Hence I would double down and argue that we did not take implicit bias seriously enough. We collectively settled for merely raising awareness about it, merely advocating reforms to circumvent it (like affirmative action, anonymous review, etc.), and merely fighting battles over the norms of explicit discourse, rather than aiming to effect more profound transformations in individuals' habits of thinking, feeling, and acting (including our own). If we had actually taken collective steps to reduce implicit biases on a somewhat broader scale, then perhaps populist demagogues could not have tapped into, rationalized, and capitalized on a critical mass of people's "gut feelings" about various "others."

Studies like Cooley and colleagues' also suggest that we might do well to continue telling an "implicit bias story" that is significantly similar to its predecessor. *Prima facie*, these studies

suggest that condemning our peers as explicit bigots could have a licensing effect, whereas insisting that their negative gut feelings and political choices “don’t reflect the real you” could do the opposite. Speaking pragmatically, I suspect we should continue trying to appeal to people’s “better natures,” making the familiar distinction between calling out “what they did” versus making claims about “who they are,” etc.

This speaks somewhat to Saul’s question #2 about what to do in reaction to the millions of people willing to tolerate discrimination, although I am, for all the familiar reasons, wary of assuming the posture of coming down from the ivory tower to transform the minds of the masses. I don’t have much to say here beyond what many others have already said. I think we can take steps to reduce our own biases (including biases against those whose political orientation differs from our own) and model the kind of behavior we’d like to see. Moreover, the kinds of intergroup integration advocated by contact theorists should be pursued across political and religious as well as racial divides. Some Trump voters complained that Clinton simply didn’t come to their town. I think there’s a real role for just showing up and listening, rather than trying to debias the political other.

Regarding #4, I don’t think it ever makes sense to prioritize the structural. The question is always *which* structures to prioritize, and there’s always a symbiotic question about which individual changes to prioritize such that the latter will promote the former. There’s a profound chicken-or-egg, Zeno-paradoxical character to this. There have to be some individuals working for the structural reforms, but then there have to be structures that enable the individuals to do so. In fact, regarding Saul’s note #5, I share her sense of difficulty in clearly and consistently distinguishing the structural from the individual. I think there is a distinction, with clearer extremes but many unclear, borderline cases. In fact, I refer readers to Section 1.2 of Saul’s most

recent draft of [“\(How\) Should We Tell Implicit Bias Stories?”](#) for an especially helpful discussion of the complex intertwining between the individual and the structural in explanations of injustice. Heuristically, I take structures to be the contexts in which individuals operate, but how we individuate contexts is going to be a messy, pragmatic affair. What counts as part of the individual versus part of the context seems up for grabs depending on our local explanatory aims. Unfortunately, I don’t have any useful suggestions about specifically which structural changes to prioritize these days. I am just as overwhelmed by the sheer number of unfolding catastrophes as everyone else. My only meek suggestions are that people focus on the stuff where they think they have the most opportunity to make a difference, and the stuff that they’re most “fired up” about.

Regarding #1, Saul raises the question whether the backlash, e.g., to Obama *qua* counterstereotypical exemplar, could have been prevented or mitigated. I’m not sure. We are not omnipotent. The racist backlash to Obama was spurred on and exploited by prominent Republicans and Fox News, who gave Trump his birther platform. It was, basically, outside our control. But if we’re entertaining counterfactuals, how far from the actual world are we allowed to go? If we had already achieved a more integrated society when the first black President was elected, then many more white people would personally know lots of counterstereotypical members of other racial, ethnic, and religious groups, and wouldn’t rely so heavily on media misinformation. And again, more serious investments could have been made in *empirically informed* approaches to prejudice reduction and diversity training. Looking at nearer possible worlds, Obama has sometimes been criticized for not being a great communicator (and not just by political opponents). I don’t know if that claim has any merit, but it’s possibly true that he could have done more in the way of “listening tours,” interacting with people face to face about

the unfolding effects of the Affordable Care Act, and so on. Alternatively, with 20/20 hindsight, perhaps we should have taken backlash to be inevitable, in which case Obama should not have tried to be a centrist or a unifying intergroup exemplar at all. The fact that he deported millions of people, expanded drone strikes and the surveillance state, and implemented a Republican-inspired, market-based healthcare expansion did nothing to prevent his being vilified as soft on immigration, weak on defense, and socialist on healthcare. Maybe he should have embraced the polarization and focused on riling up the left by fighting the good fight rather than by pursuing strategies that might appease enough centrist Republicans or independents. But the final pages of the book on Obama-*qua*-counterstereotypical-role-model have not been written. There's a whole generation of young people to whom he's been an inspiration, and we won't know the full ramifications of that influence for years to come.

References

- Anderson, E. (2012). Epistemic Justice as a Virtue of Social Institutions. *Social Epistemology*, 26(2), 163–173.
- Cooley, E., Payne, B., Loersch, C., & Lei, R. (2015). Who owns implicit attitudes? Testing a metacognitive perspective. *Personality and Social Psychology*, 41(1), 103–115.
- Cooley, E., Payne, B., & Phillips, K. (2014). Implicit bias and the illusion of conscious ill will. *Social Psychological and Personality Science*, 5(4), 500–507.
- Haslanger, S. (2012). Chapter 11, “Oppressions: Racial and Other.” In *Resisting Reality: Social Construction and Social Critique* (pp. 311–338). OUP USA.
- Kawakami, K., Dunn, E., Karmali, F., & Dovidio, J. F. (2009). Mispredicting affective and behavioral responses to racism. *Science*, 323(5911), 276–278.

